COLLEGE OF INFORMATION SCIENCES AND TECHNOLOGY
THE PENNSYLVANIA STATE UNIVERSITY

Running Behavioral Experiments with Human Participants: A Practical Guide

Frank E. Ritter, Jong W. Kim, and Jonathan H. Morgan

frank.ritter@psu.edu jongkim@psu.edu

12 December 2009
revised 20 January 2010

Phone +1 (814) 865-4453 Fax +1 (814) 865-5604

College of IST, Information Sciences and Technology Building, University Park, PA 16802

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
20 JAN 2010		**00-00-2010 to 00-00-2010**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Running Behavioral Experiments with Human Participants: A Practical Guide	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Pennsylvania State University,College of Information Sciences and Technology,Information Sciences and Technology Building,University Park,PA,16802**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

There are few resources providing practical guides on how to prepare and run experiments with human participants in a laboratory setting at colleges and universities. In our experience, we have found that undergraduate students are taught how to design experiments, and how to analyze experimental data in courses such as Design of Experiments, Statistics, etc. On the other hand, the dearth of materials available to students regarding either preparing or running experiments has led to a significant gap between theory and practice in this area, which is particularly acute outside of psychology departments. Consequently, labs frequently must not only impart these skills to students but also address misunderstandings arising from this divorce of theory and practice in their formal education. We present here a short book that can help students to run experiments effectively and more safely with human participants. In this book, our purpose is to provide hands-on knowledge and actual procedures of experiments. We hope this book will help undergraduates in psychology, engineering, and the sciences to run studies with human participants in a laboratory setting. This will particularly help students who are not in large departments, or are running participants in departments that do not have a large or long history of experimental studies of human behavior. We are generally speaking here from our background running cognitive psychology, cognitive ergonomics, and human-computer interaction studies. Because it is practical advice, we do not cover experimental design or data analyses. This practical advice will be less applicable in more distant areas but may be still of use. For example, we do not cover how to use complex machinery, such as a fMRI or ERP. We also do not cover field studies or studies that in the US require a full IRB review. This means that we do not cover how to work with unusual populations, such as prisoners, animals, and children, or how to take and use measures that include risks to the subjects or to the experimenter (e.g., saliva, blood samples, or private information). We have addressed this book toward advanced undergraduates and early graduate students starting to run experiments without previous experience; but we believe this guide will be useful to anyone who is starting to run research studies, training people to run studies, or studying the experimental process. When running an experiment, insuring its repeatability is of greatest importance?it is critical to address variations in either method or in participant behavior. Running an experiment in exactly the same way regardless of who is conducting it or where (e.g., different research teams or laboratories) is essential. In addition, reducing variance in the participants? behavior is key to an experiment?s repeatability. This book will help you achieve these requirements, increasing both your comfort and that of the participants

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**69**	

Running Behavioral Experiments with Human Participants: A Practical Guide

Frank E. Ritter, Jong W. Kim, and Jonathan H. Morgan
frank.ritter@psu.edu jongkim@psu.edu jhm5001@psu.edu
College of Information Sciences and Technology
The Pennsylvania State University
University Park, PA 16802

12 December 2009

Abstract

There are few resources providing practical guides on how to prepare and run experiments with human participants in a laboratory setting at colleges and universities. In our experience, we have found that undergraduate students are taught how to design experiments, and how to analyze experimental data in courses such as Design of Experiments, Statistics, etc. On the other hand, the dearth of materials available to students regarding either preparing or running experiments has led to a significant gap between theory and practice in this area, which is particularly acute outside of psychology departments. Consequently, labs frequently must not only impart these skills to students but also address misunderstandings arising from this divorce of theory and practice in their formal education.

We present here a short book that can help students to run experiments effectively and more safely with human participants. In this book, our purpose is to provide hands-on knowledge and actual procedures of experiments. We hope this book will help undergraduates in psychology, engineering, and the sciences to run studies with human participants in a laboratory setting. This will particularly help students who are not in large departments, or are running participants in departments that do not have a large or long history of experimental studies of human behavior.

We are generally speaking here from our background running cognitive psychology, cognitive ergonomics, and human-computer interaction studies. Because it is practical advice, we do not cover experimental design or data analyses. This practical advice will be less applicable in more distant areas, but may be still of use. For example, we do not cover how to use complex machinery, such as a fMRI or ERP. We also do not cover field studies or studies that in the US require a full IRB review. This means that we do not cover how to work with unusual populations, such as prisoners, animals, and children, or how to take and use measures that include risks to the subjects or to the experimenter (e.g., saliva, blood samples, or private information).

We have addressed this book toward advanced undergraduates and early graduate students starting to run experiments without previous experience; but we believe this guide will be useful to anyone who is starting to run research studies, training people to run studies, or studying the experimental process.

When running an experiment, insuring its repeatability is of greatest importance—it is critical to address variations in either method or in participant behavior. Running an experiment in exactly the same way regardless of who is conducting it or where (e.g., different research teams or laboratories) is essential. In addition, reducing variance in the participants' behavior is key to an experiment's repeatability. This book will help you achieve these requirements, increasing both your comfort and that of the participants who participate in your experiments.

This book consists of seven sections with eight appendices. We concisely describe below each section's contents. We hope you find it relevant and useful.

Section 1, Overview of the Research Process, describes briefly where experiments fit into the research process. If you have taken either an experimental methods course or a research design course, you can skip this chapter. If, on the other hand, you are either a new research assistant, or are working on a project in which you are unclear of your role or how to proceed, this chapter may provide some helpful context.

Section 2, Preparation for Running Experiments, describes pertinent topics for preparing to run your experiment—such as supplemental reading materials, recruitment of participants, choosing experimental measures, and getting Institutional Review Board (IRB) approval for experiments involving participants.

Section 3, Potential Ethical Problems, describes ethical considerations necessary for safely running experiments with human participants—i.e., how to ethically recruit participants, how to handle data gathered from participants, how to use that data, and how to report that data. Being vigilant and aware of these topics is a key component to rigorous, as well as ethical, research.

Section 4, Risks to Validity to Avoid While Running an Experiment, describes risks that can invalidate your experimental data. If you fail to avoid risks, you may obtain either false or uninterruptible results from your experiment. Thus, before starting your study, you should be aware of these risks and how to avoid them.

Section 5, Running a Research Study, describes practical information about what you have to do when you run the experiments. This section will give an example procedure that you can follow.

Section 6, Concluding a Research Session and Study, describes practical information about what to do at the conclusion of each experimental session.

Section 7, Example Research Studies, describes example experimental studies that give you a brief synopsis of procedural steps. You can indirectly experience a real example of an experiment. We include several examples, forms, and checklists as appendices. These forms will vary by lab and IRB, but provide examples of the style and tone.

Acknowledgements

Preparation of this manuscript was partially sponsored by a grant from the Division of Human Performance Training, and Education at the Office of Naval Research, under Contract # W911QY-07-01-0004. The views and conclusions contained in this report are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government or the Pennsylvania State University.

Ellen Bass, Richard Carlson, Karen Feigh, Alex Kirlik, and Razvan Orendovici have provided useful comments, but incompleteness and inadequacies remain the fault of the authors.

Table of Contents

1 OVERVIEW OF THE RESEARCH PROCESS 1

1.1 Overview 1
1.2 Definition of Terms 3
1.3 Further Readings 5

2 PREPARATION FOR RUNNING EXPERIMENTS 6

2.1 Literature in the Area 6
2.2 Choice of a Term: Participants or Subjects 7
2.3 Recruiting Participants 7
2.4 Subject Pools 9
2.5 Care, Control, Use, and Maintenance of Apparatus 9
2.6 Testing Facility 11
2.7 Choice of Measures: Performance, Time, Actions, Errors, Verbal Protocol Analysis, and Other Measures 12
2.7.1 Types of measures 12
2.7.2 Levels of measurement 12
2.7.3 Scales of measurement 13
2.8 Error Data 14
2.9 Run Analysis with Pilot Data 14
2.10 Institutional Review Board (IRB) 15
2.11 Further Readings 16

3 POTENTIAL ETHICAL PROBLEMS 17

3.1 Recruitment of a Broad Selection of Subjects 17
3.2 Talking with Subjects 17
3.3 Coercion of Participants 18
3.4 Sensitive Data 18
3.5 Plagiarism 19
3.6 Fraud 19
3.7 Summary 20
3.8 Further Readings 20

4 RISKS TO VALIDITY TO AVOID WHILE RUNNING AN EXPERIMENT 21

4.1 Validity Defined: Surface, Internal, and External 21
4.2 Risks to Validity 23
4.2.1 Power: How many participants? 23
4.2.2 Experimenter effects 24
4.2.3 Participant effects 25
4.2.4 Randomization 25
4.3 Example Problems 26
4.4 Further Readings 26

5 RUNNING A RESEARCH STUDY 27

5.1 Script 27
5.2 Piloting 27
5.3 Dress Code for Experimenters 27
5.4 Welcome 28
5.5 Missing Subjects 28
5.6 Decorum 28
5.7 Debriefing 28
5.8 Payments and Wrap-up 29
5.9 Simulator Studies 29

5.10 Problems and How to Deal with Them 30
5.11 Example Problems 30
6 CONCLUDING A RESEARCH SESSION AND STUDY 31
6.1 Data Care, Security, and Privacy 31
6.2 Data Backup 31
6.3 Chance for Insights 31
7 EXAMPLE RESEARCH STUDIES 33
7.1 Skill Retention Study 33
8 AFTERWORD 36
APPENDIX A: GLOSSARY 37
APPENDIX B: A CHECKLIST FOR SETTING-UP EXPERIMENTS 38
APPENDIX C: A OVERVIEW OF STEPS FOR RUNNING EXPERIMENTS 39
APPENDIX D: EXAMPLE SCRIPT TO RUN AN EXPERIMENT 40
APPENDIX E: SAFETY OF EXPERIMENTS 42
APPENDIX F: EXAMPLE CONSENT FORM 43
APPENDIX G: EXAMPLE DEBRIEF FORM 46
APPENDIX H: EXAMPLE IRB APPLICATION 47
9 REFERENCES 62

1 Overview of the Research Process

This chapter describes briefly where experiments fit into the research process. If you have had an experimental methods course or a research design course, you can skip this chapter. If you are a new research assistant, or are working on a project in which you are unclear about how you fit in, this chapter may provide some helpful context. We also define some common terms so that you can better communicate with the principal investigator and other members of your research team.

1.1 Overview

Here is a list of important topics that form our overview of the research process.

(1) Establish a hypothesis

The most common hypothesis is that one factor in the world influences another factor that can be measured. This is the most common hypothesis tested. To test this hypothesis, the first factor, such as time exposed to a stimuli has to be varied, and the second factor, such as how long something is remembered, must be measured.

Sometimes, a hypothesis can simply be that the area is interesting, and that gathering data will provide insights. This is occasionally referred to as a fishing expedition, because you do not know what you will catch. This type of hypothesis is sometimes criticized for being too general, but for exploratory work, it can be very successful. For example, we suspected that subjects when confronted with a problem would choose problem-solving strategies based upon certain key features, but we did not know which features factored into the participants' choices. So, we included multiple types of problems and multiple types of features. Then, with analysis, we were able to pull out which features participants based their decisions upon across a wide range of problem types (Reder & Ritter, 1992).

(2) Set up the experiment

To investigate what you want to know—to test your established hypothesis, it is necessary to draw a picture in your mind about what is needed to achieve your goal.

For example, suppose that you want to know how students can retain foreign vocabulary words in terms of the spacing of learning (i.e., massed or distributed practice). To set up this experiment, an experimenter would need materials for displaying the vocabulary words and for recording responses from participants (e.g., time or accuracy).

(3) Pilot study to understand the theory's edges

This stage practically helps you to identify what is working, what is not working, and what is missing for your successful investigation. Probably, a theory would explain a part of the phenomena. The theory's edge indicates a feasible region where the phenomena are scientifically explained and understood. Understanding the theory's

edge is important because you can identify a concrete direction of your research based on the understanding of the theory's edge.

For example, you might ask friends and colleagues to try a mirror tracing task. You might run people casually, in their offices, and record their times to learn how their response times differ by stimuli. You would not report these results, but use them to adjust your apparatus and your stimuli.

(4) Recruit subjects

Sometimes, recruiting subjects is easy, and then again sometimes recruiting is hard. It depends on your local circumstances. If you have access to a subject pool, it is easier. If, on the other hand, your study requires particular subjects with particular expertise (such as airplane pilots), recruiting is harder.

This step may be done while piloting and setting up the study if there are few risks to preparing the study or if subjects are hard to recruit. If subjects are easy to recruit and the study harder to prepare, it may be done in the opposite order.

(5) Running subjects

Running subjects may give you different outcomes than those of your pilot study. The primary cause for these differences is generally due to individual variability—participants may think or react in unanticipated ways. Or, you may get different results because your study is more formal. In either case or even the case where there are fewer surprises, you are interested in seeing the truth about the world.

(6) Adjust the experiment

It may be necessary to adjust the experiment to address problems identified in a previous stages, or as a second study. This process might be repeated, however eventually through this process of adjustment, the experiment will stabilize.

(7) Running a modified study

After the iteration of adjusting a series of your pilot studies, you would reach an experimental design that gives you much more stable and interpretable outcomes. It often seems that the process takes more time than you think, but this process is necessary to produce interpretable and repeatable results.

(8) Analyze results

If you have been careful, you have analyzed your pilot data to make sure that the output from the study can be analyzed. Sometimes, timestamps are in the wrong format to be read by an analysis program, or are not recorded, or the subject's name has remained attached to a file (rather than their ID). Checking that the data can be put into the analysis software is time well spent before you run participants in a larger trial.

(9) Write a final report

Take the results and prepare a manuscript: perhaps in the form of a technical report for a sponsor, as a conference paper, journal article, or maybe in a thesis. There are useful books on this area, and we do not address this topic further here.

Note that these steps are normative; they are what should happen. In practice, this process often runs in parallel, can vary in order (insights do not always come between experiments), and is iterative (Boehm & Hansen, 2001). Furthermore, breakthroughs frequently result from interactions between multiple experiments and researchers in a lab.

1.2 Definition of Terms

We list a list of terms with their definitions that are frequently used in the process of the research study.

(1) Null hypothesis vs. alternative hypothesis

The null hypothesis answers that there is nothing (null) going on in the experimental investigation—factor 1 has no influence on factor 2, and any correlation you see is just noise, a random occurrence. That is, the significance testing leads you to make a decision that there is no other influence.

The hypothesis is a competing with the null hypothesis and something that researchers want to prove to be true through the research process.

Here are some examples. One of the major findings by Ebbinghaus (1885/1913) is the principle of distributed practice. This principle holds it is desirable to spread out practice rather than massing it together in a session. That is, a 10-minute practice per day for 5 days can provide a better learning and retention rate than a 50-minute practice in one day. Based on this principle, let us construct a research hypothesis. In this case, the null hypothesis would be that there is no statistically significant difference in the learning and retention rates between distributed and massed practice groups. The hypothesis would be that a statistically significant difference does exist between the two groups, and that the distributed practice group displays higher learning and retentions rates than the massed practice group.

There may be alternative hypotheses as well. If the study is not well designed or well run, it can be that something besides the hypothesis is causing the changes observed. For example, if the first 20 subjects to show up are put into one group, and the last 20

to show up are put into another group, then the differences between groups might not be caused by how they are treated, but by whatever caused them to be in the first or last group, such as conscientiousness.

(2) Significance testing

Significance testing refers to the process used to determine whether an effect (or effects) observed in the experiment is a real effect, rather than just the result of the statistical analysis. In this significance test, you would accept or reject either the null hypothesis or the alternative hypothesis.

(3) Independent variable vs. dependent variable

For more information, please refer to Section 2.7.1.

(4) Stimuli

Stimuli are events that evoke or cause a reaction. That is, in this context, stimuli are events evoking reactions from a subject.

(5) Reaction time (RT)

Reaction time is an elapsed time between the presentation of a stimulus and the consequent response to that stimulus from the subject.

(6) Principal Investigator

A principal investigator is an official point of contact and is responsible for a grant contract. A principal investigator can also take the role of the lead researcher, responsible for conducting experiments involving human participants and gathering data.

(7) Lead researcher

A lead researcher generally indicates a person who is responsible for designing and running experiments. An experimenter actually administers experimental procedures to participants and collects data.

(8) IRB

IRB stands for an Institutional Review Board, which is a review committee to help protect the rights and welfare of human participants in a research study. An IRB can (a) approve/disapprove a research study, (b) modify a research study, (c) conduct continuing reviews, (d) observe/verify changes, (e) suspend or terminate approval, and (f) observe the consent process and the research procedures.

Studies in US institutions need IRB approval according to Federal law. Great review boards can do this quickly and help you with your research, pointing out how to be more responsible to subjects and how to improve your study (gathering data that is interpretable is also a responsibility); good review boards can do this quickly. They will also require your cooperation, and you should look forward to working with them.

(9) Informed consent

Informed consent is a process to (a) provide specific information about the research study and its procedures to the participants (or subjects), (b) answer questions to ensure the participants understand the research, (c) provide the participants with adequate time to consider their decisions, and (d) obtain the voluntary agreement from the participants to take part in the research study.

1.3 Further Readings

A course in experimental methods is probably the best way to learn about how to design and run studies. In addition, we can provide a list of suggested reading materials that provide you with further knowledge about experimental design and methods. We list them in an alphabetical order of the first author.

- Bernard, H. R. (2000). Social research methods: Qualitative and quantitative approaches. Thousand Oaks, CA: Sage.

This is a relatively large book. It covers a wide range of methods, some in more depth than others. It includes some instructions for how to perform the methods.

- Coolican, H. (2006). *Introduction to research methods in psychology* (3rd ed.). London, UK: Hodder Arnold.

- Cozby, P. C. (2004). *Methods in behavioral research* (8th ed.). New York, NY: McGraw-Hill.

- Leary, M. R. (2004). *Introduction to behavioral research methods* (4th ed.). Boston, MA: Pearson.

- Martin, D. W. (1995). *Doing psychology experiments* (4th ed.). Pacific Grove, CA: Brooks/Cole Publishing.

- Ray, W. J. (2003). *Methods: Toward a science of behavior and experience* (7th ed.). Belmont, CA: Wadsworth/Thompson Learning.

This is a book for the first course in experimental methods in psychology. It is a useful and gentle introduction to how to create and run studies and how to present the results.

2 Preparation for Running Experiments

Broadly speaking, scientists and engineers investigate scientific inquiries; and experiments constitute one powerful form of investigation. For many studies, human participation is necessary to adequately explore the question. For instance, evaluating the usability of a haptic interface (e.g., a Wii remote) before its introduction to the market would be an example. The question, then, is what considerations should inform the investigator when conducting these kinds of research studies.

In general, scientific inquiries in the areas of human-computer interaction (HCI), human factors, cognitive psychology, and cognitive science require the involvement of human participants. One distinguishing factor of these disciplines, and thus experiments in these areas, has been the centrality of the *human participant*.

Consequently, working in these areas requires not only understanding the theoretical and ethical issues incumbent to running human participants but also the practical aspects of the process itself. To start to frame this discussion, we provide an overview of this process, and issues related to it.

Let us again consider a usability study evaluating a haptic interface. For this investigation, a lead research scientist or a lead researcher would establish a study hypothesis and design an experiment by first defining: what to measure (dependent variables), what factors to manipulate (independent variables), and what environmental conditions to consider.

There is a lab. Perhaps in this lab, multiple experiments are going on at the same time. Joining the lab as a new research assistant, you have come to help out and to learn in this area, specifically with running research studies. What do you do? Where do you start? How do you avoid common and easily fixed problems?

2.1 Literature in the Area

This book does not assume that you have a background in statistics or studied experimental design, but to help run a study you often do not need to know these areas (but they do help!). If you need help in these areas, there are other materials that will prepare you to design experiments and analyze experimental data. In addition, most graduate programs with concentrations in HCI, cognitive science, or human factors engineering feature coursework that will help you become proficient in these topics.

Many introductory courses in statistics, however, focus primarily on introducing the basics of ANOVA and regression. These tools are unsuitable for many studies analyzing human subject data where the data is qualitative or sequential. Care, therefore, must be taken to design an experiment that collects the proper kinds of data. If ANOVA and regression are the only tools at your disposal, we recommend that you find a course focusing on the design of experiments featuring human participants, as well as the analysis of human data, and that you gather data that can be used in a regression because it can be used to make stronger predictions.

Returning to the topic of readings, it is generally useful to have read in the area in which you are running experiments. This reading will provide you further context for your work, including

discussions about methods, types of subjects, and pitfalls you may encounter. For example, the authors of one our favorite studies, an analysis of animal movements, notes that data collection had to be suspended after having been chased by elephants! If there are elephants in your domain, it is useful to know about them. There are, of course, less dramatic problems such as common mistakes subjects make, correlations in stimuli, self-selection biases in a subject population, power outages, printing problems, or fewer participants than expected. While there are reasons to be blind to the hypothesis being tested by the experiment (that is, you do not know what treatment or group the subject is in that you are interacting with, so that you do not implicitly or inadvertently coach the subjects to perform in the expected way), if there are elephants, good experimenters know about them, and prepared research assistants particularly want to know about them!

As a result, the reading list for any particular experiment is very individualized. You should talk to other experimenters, as well as the lead researcher about what you should read.

2.2 Choice of a Term: Participants or Subjects

Disciplines vary as to which term they prefer: *subject* or *participant.* *Participant* is the newer term, and was adopted by many research communities to emphasize the researcher's ethical obligations to those participating in their experiments. Nevertheless, *subject* is still commonly used, and appears in older research. For students in many psychology programs, the term, *participants,* is preferred to that of *subjects*. The *Publication Manual of the American Psychological Association (APA)*, 5th ed. (American Psychological Association, 2001, p. 70) suggests replacing the impersonal term, *subjects*, with the more descriptive term, *participants*. The APA goes on to define *participants* as individuals: college students, children, or respondents.

Whether following the APA guideline or not, we should recognize that *S*, *Ss*, *S's*, *E*, *Es, E's* indicate *Subject*, *Subjects, Subject's*, *Experimenter*, *Experimenters,* and *Experimenter's* in earlier research —Fitts's 1954 study is one example. Furthermore even within the discipline of psychology, opinion can be split. Roediger (2004) argues against the change to *participants* made in the latest version of the *APA's Publication Manual*. He argues that *subjects* is both more consistent and clearer, noting that the term has been in use since the 1800's and that it better defines the relationships involved. He argues that the term, *participants,* fails to adequately capture the distinction between the experimenter and those in the study—strictly speaking experimenters are participants as well. We use these terms interchangeably in this document because we recognize other research communities may still prefer *subjects*, and because not all psychologists are members of the APA.

2.3 Recruiting Participants

Recruiting participants for your experiment can be a time- consuming and potentially difficult task, but it is a very important procedure to produce meaningful data. An experimenter, thus, should carefully plan out with the lead researcher (or the principal investigator) to conduct successful participants recruitment for the research study. Ask yourself, "What are the important

characteristics that my participants need to have?" Your choices will be under scrutiny, so having a coherent reason for which participants are allowed or disallowed into your study is important.

First, it is necessary to decide a population of interest from which you would recruit participants. For example, if an experimenter wants to measure the learning effect of foreign language vocabulary, it is necessary to exclude participants who have prior knowledge of that language. In addition, it may be necessary to consider age, educational background, gender, etc., to correctly choose the target population.

Second, it is necessary to decide how many participants you would recruit. The size of participants can affect your final results. The more participants you can recruit, the more reliable your results will be. However, limited resources (e.g., time, money, etc.) force an experimenter to find the appropriate and reasonable number of participants. You may need to refer to previous studies to get some ideas of the number of participants, or may need to calculate the power of the sample size for the research study, if possible (most modern statistical books have a discussion on this, and teach you how to do this, e.g., Howell, 2008).

There are several ways that participants can be recruited. The simplest way is to use the experimenters, themselves. In simple vision studies, this is often done because the performance differences between people in these types of tasks is frequently negligible and knowing the hypothesis to be tested does not influence performance. Thus, the results remain generalizable even with a small number of participants.

The next way that subjects can be recruited that we will consider is a sample of convenience. Samples of convenience consist of people who are accessible to the researcher. Many studies use this approach, so much so that this is not often mentioned. Generally for these studies, only the sampling size and some salient characteristics are noted that might possibly influence the participants' performance on the task. These factors might include age, major, sex, education level, and factors related to the study, such as nicotine use in a smoking study, or number of math courses in a tutoring study.

In studies using samples of convenience, try distributing an invitation email to a group mailing list (e.g., students in the psychology department or an engineering department). Also, you can post recruitment flyers in a student board, or make an advertisement in a student newspaper. Use efficiently all resources and channels that are available to you.

There are disadvantages to using a sample of convenience. Perhaps the largest is that the resulting sample is less likely to lead to generalizable results. The subjects you recruit are less likely to represent a sample from a larger population. Students who are subjects are different from students who are not subjects. To name just one feature, they are more likely to take a psychology class and end up in a subject pool. And, the sample itself might have hidden variability in it. The subjects you recruit from one method (an email to them) or from another method (poster) may be different, and we know they differ over time, those that come early to fulfill a course requirement are more conscientious than those that come late. So, for sure, randomly assign these types of subjects to the conditions in your study.

The largest and most carefully organized sampling group is a random sample. In this case, researchers randomly sample a given population by carefully applying sampling methodologies meant to ensure statistical validity and equal likelihood of selecting each potential subject. Asking students questions at a football game as they go in does not constitute a random sample—some students do not go (selection bias). Other methods such as selecting every 10th student based on a telephone number or ID introduce their own biases. For example, some students do not have a publicly available phone number, and some subpopulations register early to for their ID numbers. Truly choosing a random sample is difficult, and you should discuss how best to do this with your lead researcher.

2.4 Subject Pools

One approach for recruiting participants is a *subject pool*. Subject pools are generally groups of undergraduates who are interested in learning about psychology through participation. Most Psychology departments organize and sponsor subject pools. For more information, refer to the next section.

Subject pools offer a potential source of participants. You should discuss this as an option with your lead researcher, and where appropriate, learn how to fill out the requisite forms. If the students in the study are participating for credit, you need to be particularly careful with recording who participated because the students' participation and the proof of that participation represent part of their grade.

A whole book could be written about subject pools. Subject pools are arrangements that psychology or other departments provide to assist researchers and students. The department sets up a way for experimenters to recruit subjects for studies. Students taking particular classes are either provided credit towards the class requirement or extra credit.

The theory is that participating in a study provides additional knowledge about how studies are run, and provides the participant with additional knowledge about a particular study. The researchers, in turn, receive access to a pool of potential subjects.

When students do not wish to participate in a study, alternative approaches for obtaining course credit are provided.

2.5 Care, Control, Use, and Maintenance of Apparatus

What materials do you need to run experiments? The experiments in a controlled environment (e.g., a laboratory) usually require participants to interact with a computer device, a prototype, or a mock-up. For example, it is possible to implement a task environment in a computer screen—such as an air traffic control task like Argus (Schoelles & Gray, 2001), a driving simulator like Distract-R (Salvucci, in press), experimental tasks with E-Prime (e.g., MacWhinney, St. James, Schunn, Li, & Schneider, 2001), or a spreadsheet task environment (Kim, Koubek, & Ritter, 2007). Part of what you will have to do is to understand the task environment so that you can prepare it for each session, save the data if it collects data, and shut it down after each session.

As you begin to work on your research task, you are likely to consider several approaches for improving your study. Finding, developing, or modifying the task environment to support your study is often an early consideration. The task environment provides the setting for investigating the questions of interest, and having the right task environment is a key element to a successful study. If designing and implementing a new task environment for your research study seems infeasible, try reusable and sharable environments.

After choosing and setting up the task environment, the next step is to determine what method you will use to record user performance. Data collection deserves serious thought, in an attempt to provide meaningful results. Data can be qualitative (i.e., not in a numerical form) or quantitative (i.e., in a numerical form). Different hypothesis and theories require different types of data to test them, and thus methods to collect data. For example, you can use a camcorder in an interview to gather qualitative information or a keystroke logger like RUI (Kukreja, Stevenson, & Ritter, 2006) to measure numerical values of quantitative data in unobtrusive and automatic ways. We suggest avoiding manually recording data—it is hard, takes a significant amount of time, and is prone to error. Though, sometimes, manual data collection is unavoidable; often with a little forethought ways can be found to automate the process.

An apparatus is often required to gather behavioral data. In cognitive science, recording user behavior by using experimental software, a video recorder, a voice recorder, or a keystroke/mouse logger, etc are all common practices. There are also tools for generating studies such as ePrime. Also, some studies require using an eye-tracker to gather eye-movement data.

Experimental software

Many studies are performed with custom built, or bespoke software. The research team conducting the study usually develops these custom applications; and they can vary from a simple program to present stimuli and record reaction times to more complex programs (interactive simulations for instance). As new research assistant, you will be instructed on how to start up and run the software necessary for your work. On the other hand, as you run subjects with such programs, try moving from a passive to an active user. Make any suggestions that you think might improve the program's usability as they arise, note mistakes in the program, and observe how subjects interact with the program in novel or interesting ways. These insights can lead to further studies and to further hypotheses to test.

E-Prime

E-Prime[1] was the first commercial tool designed to generate psychological experiments on a personal computer (MacWhinney, St. James, Schunn, Li, & Schneider, 2001). E-Prime is compatible with Microsoft Windows® XP/Vista. PsyScope[2] is another experiment generation program, and a predecessor of E-Prime. You can download it free under a GNU General Public License[3]. PsyScope runs on the Macintosh. You may be asked to use these tools in your current study or may find them to be great value in producing study stimuli more quickly.

[1] http://www.pstnet.com/products/e-prime
[2] http://psy.ck.sissa.it
[3] http://www.gnu.org/copyleft/gpl.html

Keystroke loggers

It is often useful to record the user's behavior while they perform the task, not just the total task time. This can be done in several ways. Some researchers have used video recordings. This provides a very stable result that can include multiple details. It also can provide a rich context, particularly if both the subject and his or her surroundings are recorded. On the other hand, analyzing video recordings is time consuming and can be error prone. Analyzing video data requires examining the video frame-by-frame to find when the user performs each action, and then recording each action by hand into your dataset.

Another approach is to record just the keystrokes or mouse clicks. There are commercial versions available from companies like Noldus that will record keystrokes. We have also designed a keystroke logger, RUI (Recording User Input). RUI is a keystroke and mouse action logger for the Windows and Mac OS X platforms (Kukreja, Stevenson, & Ritter, 2006). It is very useful tool for recording user behavior in human-computer interaction studies. RUI can be used to measure response times of participants interacting with a computer interface over time.

Using RUI, however, does raise issues regarding privacy in public clusters (e.g., a classroom). University policies almost universally prohibit installing any tool for experimentation that obtains a user's information on identity such as a login ID or a password (Kim & Ritter, 2007). Fortunately, Kim and Ritter (2007) describe one possible portable solution to this problem. They used a simple shell script to automatically run RUI on an external drive, a jump drive. Because RUI is operated from an external drive it provides a way to efficiently use RUI on public cluster machines and then remove it when the study is over.

Eye-trackers

An eye tracker is a device to measure eye positions and movements. It can offer useful data of cognitive processes when a user interacts with an interface (e.g., a computer screen, a physical product, etc). This device is sensitive, requiring special care to guarantee the measurement's quality.

2.6 Testing Facility

A testing facility can be called a psychological testing room, human factors lab, an ergonomics lab, a usability lab, or a HCI lab. Rosson and Carroll (2002) state that a usability lab is a specially constructed observation room. In this observation room, an investigator can simulate a task environment and record the behavior of participants. Thus, the room should be insulated from outside influences, particularly noise. However, it is sometimes necessary to observe and record behaviors of a group of participants interacting with each other. In these cases, it may be hard to capture this data in a lab setting. Ideally, the testing facility should be flexible enough to conduct various types of human involved research.

Jacob Nielson (1994) edited a special issue about usability laboratories. This special issue provides several representative usability laboratories in computer, telecommunications, and consumer product companies (e.g., IBM, Symantec, SAP, Phillips, or Microsoft, etc.). You can obtain more details about these facilities from this special issue. While this special issue is

somewhat dated, the underlying concerns and some of the technological details remain accurate, in addition many of the social processes and uses for video have only become more important.

If you are designing your own study, you should try to arrange access to a room that allows participants to focus on the experimental task. Lead researchers will often have such rooms, or can arrange access to them.

2.7 Choice of Measures: Performance, Time, Actions, Errors, Verbal Protocol Analysis, and Other Measures

2.7.1 Types of measures

There are several types of measures. Questionnaires are one common and flexible type. By answering the questions, participants self-report about the question, thus providing researchers insights into their behavior. The quality and type of these responses, however, depend upon the quality and type of the questions asked—so carefully selected and carefully worded questions are important.

One example where questionnaires can be used effectively is studying self-judgment and it effects. Under certain conditions, our feelings about our knowledge and our actual knowledge may differ. In this case, our hypothetical researcher asks participants to make a judgment about what they know after memorizing vocabulary words. Using a Likert scale is a common way to measure self-judgment. Likert scales typically consist of five points with questions ranging from "Strongly disagree" to "Strongly agree". Our hypothetical researcher would then test the participants and compares the participants' responses about their knowledge with the results.

Other types of measures can include physiological measures. Cozby (2004) introduces a few popular physiological measures such as galvanic skin response (GSR), electromyogram (EMG), and electroencephalogram (EEG) that help us understand psychological variables. Also, fMRI (functional magnetic resonance image) is a popular method of measuring and examining brain activities. If you are interested in learning more about these techniques, refer to the section of Further Readings, specifically *Psychophysiological recording* (Stern, Ray, & Quigley, 2001).

2.7.2 Levels of measurement

Often within a single study, multiple measures with different characteristics are gathered. For instance, you can measure the task completion time; or you can measure the number and the times of the keystrokes and mouse actions performed by the participants during the task. You can also measure what errors were made during the task, and so on. Let us discuss some common measures taken in an HCI or cognitive science experiment.

It is necessary to decide what you are observing and measuring from the participants who are performing the experimental task. The decision is important because the choice of measures is directly related to what aspects of the participants' behavior is being captured by the task. In general there are two types of variables: (a) independent variables, and (b) dependent variables.

Independent variables cause, or manipulate, the changes in the participants' behavior that the researchers seek to observe during the study. Thus, independent variables are sometimes called manipulated variables, treatment variables, or factors (Keppel & Wickens, 2004).

To cement our understanding of variables, let us presume that we want to measure how humans forget something they have learned. We will return to this example in more detail in a later chapter; but for now, we will focus the study's independent and dependent variables. Variables that can manipulate forgetting performance include training types, retention intervals (how long a participant will retain learned information), or input modalities (what types of skills a participant is to learn). Thus, we would consider these variables the study's independent variables. They deliberately vary to create the effects, they are independent.

Dependent variables indicate what we will observe. Their values are (presumed to be) dependent on the situation set up by the independent variables. Dependent variables can either be directly observed or may be derived. The NASA TLX, for example, allows researchers to derive a measure of workload. We directly measure each of the six individual subscales, but the results from the tradeoff questions are derived. In fact, performance is often a derived measure. That is, the dependent variable is affected by the manipulation of the independent variable. We can observe the time that is required to complete a task if the investigation is to understand human performance caused by forgetting. Also, we can observe errors produced by participants to measure forgetting. These variables are considered to be dependent variables. There can be one or more dependent variables. One dependent variable in an experiment is referred to univariate methods, and more than two dependent variables are referred to multivariate methods.

To sum up, dependent variables are the responses being observed during the study while independent variables are those factors that researchers manipulate to either cause or change those responses.

2.7.3 Scales of measurement

Variables can be basically measured using four scales (Ray, 2003): (a) nominal measurements, (b) ordinal measurements, (c) interval measurements, and (d) ratio measurements. Knowing these scales of measurement is important because the data interpretation techniques available to you for interpreting the results are a function of the scales of measurement used, and the use of such data, perhaps even how it is stored depends on what kind of data it is.

Nominal (also referred to as categorical) measurements are used to classify or name variables. There is no numeric measure of values representing names or separate categories. For example, participants can be classified into two groups—a male group and a female group, to measure performance on using a GPS navigation system. In this case, the gender difference is an independent variable to compare performance. Or, if the numbers 1 to 10 are treated as words, such as how often they are said, then there is not necessarily even an order to them, they could be sorted alphabetically.

Ordinal measurements, in contrast, represent some degree of quantitative difference (or relative amount). For example, football rankings in the Big Ten conference are an ordinal measurement,

as are ratings on a scale of 1 to 10. Differences between the first and second team, between 9th and 10th, and between ratings of 4 and 5 and 6 and 7 are not necessarily equal, just ordered.

Interval measurements rely upon a scale values based on a single underlying quantitative dimension. The distance, therefore, between the consecutive scale values are meaningful. For example, the interval between 6 and 12 equals the interval between 12 and 18. That is, the distance between the consecutive values is 6.

Ratio measurements determine value with respect to an absolute zero—there is no length shorter than 0 inches for instance. The most common ratio measurement can be found in a count measure (i.e., the number of hits or misses). For example, in a shooting game, the number of hits is used to determine the firer's accuracy.

Frequently, sets of sequential data, or protocols, are gathered from human participants for a given task. Protocols may be multiple streams of data including verbal utterances, motor actions, environmental responses, or eye movements (Newell & Simon, 1972). As an example of a verbal protocol, consult the testing methodology developed by Ritter and Larkin (1994) for the principled analysis of user behavior.

Verbal data often provides insights into understanding human behavior. Ericsson and Simon (1993) published a summary of how and when to use verbal reports as data to observe humans' internal cognitive processes. The basic assumption of the verbal protocol theory is that verbalization of a human's memory contents (not their view of their thought processes) can be used to derive the sequence of thoughts to complete a task. Thus, verbalization can be a valid form of data representation that offers us certain unique insights into cognition (see chunking). This type of data requires audio recordings, and often comes with special apparatus for recoding and special software and tools for analyzing the results. It is time consuming, but can be very helpful for understanding how the task is performed.

2.8 Error Data

Another type of data to gather is error data. Error data consists of trials or examples where subjects did not perform the experimental task or some aspects of the task correctly. This type of data can provide useful examples of where cognition breaks down. In addition, it helps describe the limits of performance and cognition.

Error data is generally more expensive to collect because in most cases participants perform the task correctly. Thus, more trials have to be run to gather a hundred errors than it takes to gather a hundred correct responses. If errors are not interesting theoretically for your research study, some pilot running of the experiments may be required to generate an experiment where errors do not occur too often.

2.9 Run Analysis with Pilot Data

Before launching your experimental study, we can highly recommended that you run a pilot subjects, gather data from them, and analyze the data. The number to run can be found with experience, or by talking with your PI. Analysis of pilot data can provide a baseline, or identify

problems with the testing techniques or measures used. Your pilot subjects can be your friends, family, or subjects taken from your subject pool.

If the results from the pilot data are not what you expected, you can revise the design of experiments (e.g., change independent variable, change the target task, or add another treatments, etc.), keeping in mind that the answer might be that your assumptions are wrong and that using a small number of subjects only allows you to see large effects. Then, you will need to gather more pilot data. If the results from the pilot data match your expectations, plan to launch your experiments to gather data. If not, an interesting new study topic may have emerged.

2.10 Institutional Review Board (IRB)[4]

Investigators in psychology or human factors must obtain approval from the appropriate host institutions or organizations prior to conducting research. The organization charged with approving research applications in a university setting is called the *Institutional Review Board* (IRB). The IRB is a committee monitoring, approving, and reviewing biomedical and behavioral research involving humans. To protect the rights of research participants, universities have established an *Institutional Review Board* (IRB).

Before the onset of the experiment, investigators must obtain the informed and voluntary consent of the participants selected for the study. The American Psychological Association's Ethical Principles of Psychologists and Code of Conduct[5] specifies that participants have the right to informed consent—participants have the right to understand what will happen in the study (e.g., any known risks of harm, possible benefits, and other details of the experiment). Only after receiving such a briefing, can a participant agree to participate in the experiment. Thus, the details of the experiment should be written in clear jargon free language, and without any reference to special technical terms. The participants must be able to easily understand the informed consent form. In addition, the form should enable prospective participants to determine, for themselves, whether they are willing to participate given his or her situation and personal tolerance for risk. We provide an example of an informed consent form in the Appendix.

There are a few exceptions that are worth noting, where IRB approval is not required. If you are running yourself and only yourself, you do not need IRB approval. If you are running studies only for class work, programmatic improvement and not for publication, then IRB is not required. These exceptions are useful when you are piloting studies, or when you are teaching (or learning). Of course, you can in most cases still seek IRB approval in these cases. The approval process offers you the opportunity for feedback on how to make your study more safe and efficient. Approval also allows later publication if the results are interesting.

IRB approval is required before any aspect of the study that will be published is conducted, including subject recruitment. Without exception, IRB approval cannot be granted once the study has been conducted. Consequently, you should seek IRB approval early in the process and keep your timeline and participant count as "loose" as possible. You do not need to seek new

[4] This applies to research in the US. You should enquire locally because some countries do not see risk in routine cognitive experimental projects.
[5] http://www.apa.org/ethics/code2002.html

approval for enrolling fewer participants than requested or finishing early. You will, however, need to seek approval for running behind or for enrolling a larger number of participants.

IRB policies are subject to interpretation so when in doubt contact the IRB representative at your institution.

In general IRB reviews fall under two categories, expedited or full review. Most behavioral science studies that do not involve the use of experimental drugs, radiation, or medical procedures can be considered for expedited review. Expedited review does not require full IRB approval, and can usually be accomplished within a few weeks (again this will vary by institution). For all other cases, you will need to go through a full review—these are usually scheduled far in advance at some specified interval.

2.11 Further Readings

We list some reading materials that will help you plan and run experiments, and report results from the experiment.

- Rosson, M. B., & Carroll, J. M. (2002). *Usability engineering: Scenario-based development of human-computer interaction*. San Francisco, CA: Morgan Kaufmann Publishers.

This book provides comprehensive background of the area of human-computer interaction.

- Stern, R. M., Ray, W. J., & Quigley, K. S. (2001). *Psychophysiological recording* (2nd ed.). New York, NY: Oxford University Press.

Psychophysiological Recording is a very useful book for anyone who conducts experiments with human participants to measure their psychological or physiological responses. The book provides not only practical information regarding recording techniques but also the scientific contexts of the techniques.

- Nielsen, J. (ed.) (1994). Special issue: Usability laboratories. *Behaviour & Information Technology, 13*(1-2).

This is a specially edited article concerning usability laboratories. This special issue provides several representative usability laboratories—mostly computer, telecommunications, and consumer product companies (e.g., IBM, Symantec, SAP, Phillips, or Microsoft, etc.).

- Ray, W. J., & Slobounov, S. (2006). Fundamentals of EEG methodology in concussion research. In S. M. Slobounov & W. J. Sebastianelli (Eds.), *Foundations of sport-related brain injuries* (pp. 221-240). New York, NY: Springer.

This book chapter provides you with background for using EEG and its processes, including physiological basis and frequency analysis of the EEG. In addition, Ray and Slobounov explain EEG research on motor processes in general and brain trauma.

3 Potential Ethical Problems

There are several topics that you need to keep in mind when running subjects. Chief among these are the ethics pertaining to the running of participants, and the gathering and reporting of data including published and unpublished documents. If you have any questions, you should contact the lead researcher (or principal investigator), or other resources at your university.

3.1 Recruitment of a Broad Selection of Subjects

The results we find we would like to generalize to a wide population, indeed, the whole population. It is useful to recruit a representative population of subjects to accomplish this. It has been noted by some observers that experimenters do not always recruit from the whole population. In some studies, this is a justifiable approach to ensure reliability (for example, using a single sex in a hormonal study) or to protect subjects who are at greater risk because of the study (for example, non-caffeine users in a caffeine study).

Where there are not threats to validity, experimenters should take some care to include a representative population. This may mean putting up posters outside your department, and it may include paying attention to sex balance and even age balance in a study, and, then correcting the balance by recruiting more subjects with these features.

As the research assistant, you can be the first to notice this, and to bring it to the attention of the investigator.

3.2 Talking with Subjects

When you first welcome the subjects to your study and the study area, you might feel uncomfortable. After you have run a few sessions, this discomfort will go away. In a simple study, you can be quite natural, as there is nothing to 'give-away'. In more complex studies, you will be busy setting up the apparatus, and this tends to make things easier.

In nearly all cases, abstaining from extraneous comment on the study is an important and useful practice that makes all parties concerned more comfortable. Many experimental protocols require not giving the subject feedback during the study. In these cases, you should inform the participants at the beginning of the session that you are not allowed to provide them feedback on their performance. Generally, the debriefing can handle most questions, but if you are not sure how to answer a question, either find and ask the investigator, or, take contact details from the subject and tell them you will get them an answer. And then, do it! This also means that when you are running subjects for the first couple of times that someone who can answer your questions should be available.

In social psychology studies or where deception is involved, you will be briefed by the investigator and will practice before hand. In this area, practice and taking advice from the main investigator is important.

3.3 Coercion of Participants

Coercion is an ethical violation of the rights of human participants. It is necessary to avoid any procedures in a study that restrict participants' freedom of consent regarding their participation in a study. Some participants, including minors, patients, prisoners, and individuals who are cognitively impaired are more vulnerable to coercion. For example, enticed by the possibility of payments, minors might ask to participate in a study. If, however, they do so without parental consent, this is unethical because they are not old enough to give their consent—agreements by a minor are not legally binding.

Students are also vulnerable to exploitation. The grade economy presents difficulties, particularly for course where a lab component is integrated into the curriculum. In these cases, professors must not only offer an experiment relevant to the students' coursework but also offer alternatives to participating in the experiment.

To address these problems, it is necessary to identify any potential condition that would compromise the participants' freedom of choice. For instance, in the second example, recall that it was necessary for the professor to provide an alternative way to obtain credit. In addition, this means ensuring that no other form of social coercion has influenced the participants' choice to engage in the study. Teasing, taunts, jokes, inappropriate comments, or implicit quid pro quo arrangements are all inappropriate. These interactions can lead to hard feelings (that's why they are ethical problems!), and loss of good will towards experiments in general and you and your lab in particular.

3.4 Sensitive Data

When preparing to run the study, you should prepare how to deal with sensitive data. There are at least two issues here—data that you anticipate is sensitive and unexpected data that arises that is sensitive.

Data that is intrinsically sensitive should be handled carefully. Personal data is the most common. Information on an individual, such as related to race, creed, gender, gender preference, religion, friendships, and so on, must be protected. This data should not be lost or mislaid. It should not be shared with people not working on the project, either formally if you have an IRB that requires notice, or informally, if your IRB does not have this provision (this may occur more often outside of the US). You should seek advice from your colleagues about what practices are appropriate in your specific context. In some situations, you are not allowed to take data from the building, and in most cases, you are encouraged to back it up and keep the backed-up copy in another safe location.

The second type of sensitive data is data that can arise where the subject's responses have implications outside of the scope of the study. This can include subjects implicating themselves in illegal activity, or unintentionally disclosing an otherwise hidden medical condition. For example, if you are administering caffeine, and you ask the subject what drugs they take (to avoid known caffeine agonists or antagonists), you may find information about illegal drug use. If you take subject's heart rate or blood pressure measurements, you may discover symptoms of underlying disease.

Generally, preparation for a study should involve discussions about how to handle sensitive data, and if there is a chance that the study may reveal sensitive data about the participants. You should fully understand how your institutions policies regarding sensitive data, and how to work with the subjects when sensitive information becomes an issue. If you have questions, you should ask the principle investigator.

3.5 Plagiarism

Plagiarism refers to taking other's work or ideas and using them as one's own, that is, without attribution. Particularly in academia, this problem is taken seriously.

An individual might be tempted to steal others' ideas, research methods, or results from unpublished or published works. Nowadays, manuscripts that are about to be submitted or already submitted for review, can be available online.

Why people are tempted to plagiarize others' work? Generally, pressure to meet or surpass institutional standards causes people to plagiarize. To pass a programming class, students might copy another student's code. A faculty member, facing review for tenure and stressed by the number of his or her refereed publications, or an RA trying to fill in a methods section all might be tempted to steal the work of others. Sometimes, the pressure to publish, is enough to tempt an academic to plagiarize other's ideas and fabricate their data.

The integrity and development of scientific knowledge is rooted in the proper attribution of credit. In the APA's publication manual (p. 349), you can find the APA's guidelines for giving credit. Direct quotes require quotation marks and citations while paraphrasing or in anyway borrowing from the work of others requires a citation. You may also need to acknowledge people who give you unpublished ideas for your research designs. In particular, you may have personal communications (e.g., email, messages from discussion groups on the net, letters, memos, etc.) that require acknowledgement. In this case, you will need to remember who gave you the idea (an email thanking them can be a good way to document this), and then cite them in the text with a date.

3.6 Fraud

We, sometimes, are shocked by news about research fraud. For example, if a researcher fabricates data and publishes a paper with the data, this is fraud. Other scientists trying to replicate the results are often the ones who find and reveal the initial findings to be fraudulent. While research fraud is unusual, we, nevertheless, must be aware that fraud can cause significant adverse effects on not only for the perpetrator of the fraud but also often second or third parties such as his or her academic institution, funding agency, or corresponding journal editor. Or, more distant people who base an educational system on a learning theory, or teaching strategies on incorrect data on memory.

If data is lost, it is lost, do not replace it. If you delete data, do not replace it. If you did not run a subject, do not run yourself. All of these practices undermine your study's validity and are extremely egregious ethical violations. It is sad when you read in an article that "data from 3 subjects were lost", but it is far better to write this than to commit fraud.

3.7 Summary

This chapter notes a few of the most important ethical problems you might face. You may encounter others. If you have questions, you should contact the lead investigator or other senior personnel. In some cases, as in many ethical situations, there may not be a right answer, there may be several right answers, and often there are better answers and good, accepted practices.

3.8 Further Readings

Here is a list of further readings for you concerning this chapter.

- You should refer to the APA's webpage, *Ethical Principles of Psychologists and Code of Conduct*. The first version was published in 1992, but has been superceded by a newer release issued in June 2003. Here is the link for the current code of conduct: http://www.apa.org/ethics/code2002.html
- American Psychological Association. (2001). *Publication manual of the American Psychological Association*. Washington, DC: American Psychological Association.

  The APA publication manual provides useful guidance for reporting your experimental findings in a written paper.

4 Risks to Validity to Avoid While Running an Experiment

Understanding how subjects will complete the task and working towards uniformity across all iterations of the task are important. The repeatability of the experiment is a necessary condition for scientific validity. There are, however, several well known effects that can affect the experimental process. Chief among these is the experimenter's effect, or the influence of the experimenter's presence on the participants. Depending upon the experimental context, the experimenter effect can lead to either better or decreased performance. The magnitude and type of effect that can be attributed to this effect generally depends upon the type and extent of personal interaction between the participant and experimenter. Thus, you should strive to provide each participant a comfortable but neutral testing experience.

Besides the experimenter effect, there are other risks to the experimental process. We hope here to not only highlight some but also illustrate how to avoid them, either directly or through proper randomization. Randomization is particularly important because you will most likely be responsible for implementing treatments while understanding the other risks will help you take steps to minimize them. Finally, there are other experimental effects that are outside of your control—we do not cover these here. Generally, these effects are associated with some idiosyncrasy (usually an event) in the testing environment. Even though you cannot eliminate all contingent events, you can note idiosyncrasies and with the principle investigator either correct or report them for future trials.

Another common source of variation across trials is the effect of the experimental equipment. For instance, if you are having subjects interact with a computer or other fixed display, you should take modest steps to make sure that the participant's distance to the display is the same for each subject—this does not mean, necessarily, putting up a tape measure, but in some cases, it does. It is necessary to be aware that the viewing distance can affect more blurred vision, irritated eyes, headache, and movement of torso and head (e.g., Rempel, Willms, Anshel, Jaschinski, & Sheedy, 2007). The factors of which can, thus, be risks to validity. Furthermore, if subjects are picking up blocks or cards or other objects, the objects should either always be in the same positions, or they should be always randomly placed because some layouts of puzzles can make the puzzles much easier to solve. The experimental set up should not be sometimes one and sometimes the other.

There will be other effects where variation in the apparatus can lead to unintended differences, and you should take advice locally to learn how to reduce them.

4.1 Validity Defined: Surface, Internal, and External

We refer to validity as the degree to which an experiment leads to an intended conclusion from the data. In general, two types of validity, internal validity and external validity, are of interest. Internal validity refers to how well experimental treatments explain the outcomes from the experiment. The experimental treatments indicate independent variables that you design. External validity, in contrast, refers to how well the outcomes from the experiment explain the phenomena outside the designed experiment. This is known as "generalizability".

Campbell and Stanley (1963) discusses 12 factors that endanger the internal and external validity. We need to consider how to reduce or eliminate the effects from these factors to guarantee valid results.

Regarding internal validity, when you run studies you may notice these factors. Good principle investigators will appreciate you bringing them to their attention. You should not panic, some of these are inevitable in some study formats, but if they are unanticipated, then they may be interesting or the study may need to be modified to avoid them.

- History: Besides the experimental variable, a specific event could occur between the first and second measurement. Typically, this is some news item such as a space launch or a disaster that influences subjects in a global way leading to better or worse results than would occur at other times.
- Maturation: Participants can grow older, become hungrier, or become more tired with the passage of the time. Thus, if you measure students at the beginning of the school year and then months later, they may get better scores based on having taken classes.
- Testing: The effects of taking a test on the scores of a second test. Thus, if you take an IQ test, the same test, a second time. You are likely to score better, particularly if you got feedback from the first taking.
- Instrumentation: It is required to calibrate a measuring instrument regularly. Some instruments need to be recalibrated with changes in humidity. Failure to recalibrate can affect an experiment's results.
- Statistical regression: We need to avoid selecting groups on the basis of their extreme scores. If you select subjects based on a high score, some of those high scores will most likely not reflect the participant's normal performance.
- Biases: Differential selection of participants for the comparison groups should be avoided. Subjects that come early in the semester to get paid or credit are different from the subjects who put it off until the last week of the semester.
- Experimental mortality: There could be a differential loss of participants from the comparison groups. Some conditions could be hard on the subjects, and thus lead them to come back less.
- Selection-maturation interaction: Given that there are two groups. Participants in one group develop faster than participants in the other group. In this case, selecting one of two groups that are maturing at different rates concerning the outcome can cause the posttest differences. This factor is an interaction effect that can exist between the subject-related variable (e.g., age) and a time-related variable.

Regarding external validity, you can also notice these sometimes.

- The reactive or interaction effect of testing: A pretest could affect (increase or decrease) the participants' sensitivity or responsiveness to the experimental variable. Some pre-tests disclose what the study is designed to study. If the pre-test asks about time spent studying math and playing math games, you can bet that mathematical reasoning is being studied in the experiment.
- The interaction effects of selection biases and the experimental variable: It is necessary to acknowledge that independent variables can interact with subjects that were selected

from a population. In this case, the outcome or findings from the experiment may not be generalized to summarize a larger population.

- Reactive effects of experimental arrangements: An experimental situation itself can affect the outcome that cannot be generalized. That is, the outcome can be a reaction from the specific experimental situation.
- Multiple-treatment interference: If multiple-treatments should be applied to the same participant, the participant's performance would then not be valid because of the accumulated effects from those multiple treatments. For example, if you have learned sample material one way, it is hard to tell if later learning is the result of the new learning method presented second, or the result of the first method, or the combination of the two.

Why mention these in a book on how to run subjects? Why not just let these be mentioned in experimental design? We mention them here because if you are new RA, you may not have had an experimental design class. And yet, many of these effects will only be or mostly be visible to the person running the study. If there is an event in a country where you are running subjects like an election, and you will be comparing results to a different country where the PI is located, it is the RA that has the best chance of noticing that something unusual that is a threat to validity has happened in the study, whereas every one can notice the global result.

4.2 Risks to Validity

4.2.1 Power: How many participants?

Performance is noisy. Differences that appear could be due to a theoretical manipulation, or it could be due to chance. We now discuss the power of a statistical test, and how a test's power can influence its effectiveness. Calculating the test's power can help maximize the benefits of a set of experimental runs by helping you decide how many subjects to run. For instance while relatively rare, running too many subjects can be wasteful when the effect size is known to be large. There are other issues that investigators need to consider, such as participants' effects or experimenters' effects. We will take these issues up in the following section.

Testing a hypothesis produces two outcomes: (a) one outcome can be rejecting the null hypothesis (H_0), while the other outcome (b) can be not rejecting the null hypothesis—that is accepting the alternative hypothesis (H_a). When investigators decide to either accept or reject the alternative hypothesis, they can make two types of errors, known as Type I and Type II errors. Table 2.1 describes these errors.

Table 2.1. Type I and II error in testing the null (H_0) and experimental (H_a) hypotheses.

Decision Made	True State	
	H_0 is true	H_a is true
Reject H_0	Type I error (report a result, but no effect)	Correct decision
Fail to reject H_0	Correct decision	Type II error (report no result, but there is an effect)

In fact, if the null hypothesis (H_0) is true, investigators should fail to reject the null hypothesis. When the null hypothesis is incorrectly rejected the null hypothesis, Type I errors occur. The probability of making a Type I error is denoted by α. On the other hand, if the alternative hypothesis (H_a) is true, in fact, investigators should accept the alternative hypothesis. When the alternative hypothesis is incorrectly rejected, Type II errors occur. The probability of making a Type II error is denoted by β.

The Power of a test is defined as the probability of correctly rejecting the null hypothesis (H_0) when it is in fact true—this is denoted by $1-\beta$. In a practical sense, via the calculation of the Power, investigators are able to make a statically supported argument that there is a statically significant difference when such a difference truly exists.

4.2.2 Experimenter effects

When two or more experimenters are running the same experiment, effects or biases from experimenters can exist. Prevent possible experimenter effects is necessary for guaranteeing the validity of the experiment. Mitchell and Jolley (2007) state reasonable causes for error that investigators should avoid: (a) the loose-protocol effect, (b) the failure-to-follow-protocol effect, and (c) the researcher-expectancy effect.

First, to avoid the loose-protocol effect, when you run the experiment by different experimenters, it is necessary to write a document that describes the procedures in detail and specifies exactly when to use each with each subject. The protocol document should allow other experimenters to run the experiment in exactly the same way, providing a standardized way to run the trials. Once you finished a draft of the protocol document, you should test it with practice participants. Producing the final protocol document will require a few iterations of writing and then testing the protocols with practice participants.

The second cause of error results from an experimenter's failure to follow the experiment's protocols. There might be several reasons for not following the protocol—the reasons can include a lack of motivation to follow the protocol, or ignorance of the protocol, etc.

The third cause for error arises from the influence of the experimenter's expectations upon his or her interactions with the participants. For instance, I might be biased (consciously or unconsciously) in how I run the experiment if I know I am testing my hypothesis. After all, I have a personal incentive to reject the null hypothesis in this case. Therefore, it is preferable when possible that the experimenters interacting with the subjects be unaware of the hypothesis being tested. When this happens, it is called a double-blind study (American Psychological Association, 2001). An example would be when the RA does not know which amount of caffeine a subject received, or what condition the subject is in.

Following consistent written protocols in and unrushed manner is one way to avoid many of these errors. Please be patient and give the participants enough time to complete each procedure to best of their ability.

4.2.3 Participant effects

Participant responses and personal characteristics can also invalidate experimental results. Measurement strategies are one common cause of error. Obtrusive measurements affect participant performance, thus invalidating the results. Suppose for example that a participant is working on a task on a computer screen. The participant is told that the task completion time is measured. A running stopwatch is placed beside the participant. In this case, the means of measuring the task completion time is not appropriate because of obtrusiveness of the stopwatch. Thus, it is generally recommend that measuring participants' performance should be conducted in unobtrusive ways.

Because personal characteristics and histories influence performance, it is important to try to methodically achieve a representative sample when selecting participants. Factors such as ethnicity, gender, age, experience, native language, or working memory capacity, etc can all affect performance. Random assignment of subjects to conditions generally helps mitigate this effect. Random assignment, however, can go wrong (or be done incorrectly), or result in a suboptimal distribution. RAs often are the earliest, best, and often the only way to discover these problems.

4.2.4 Randomization

Randomization describes the process of randomly determining both the allocation of the experimental material and the order in which individual trials are to be performed (Montgomery, 2001). Random sampling is a method for selecting the entire sample group. Ray (2003) states that one way to achieve external validity is to have the participants in the experiment constitute a representative sample of the entire population. In fact, it is very hard to accomplish random sampling. After randomly selecting your participants from the population, you should randomly assign them to their experimental groups.

Statistical methods require that the observations should be independently distributed random variables. Proper randomization of the experiment makes the assumption that the independent distribution of observed data is valid and allows us to statistically analyze the behavioral data. Randomization also can be useful to alleviate bias of selecting participants.

In some situations, Montgomery (2001) states that it is difficult to randomize the experiment because of a hard-to-change variable (e.g., temperature in a chemical process, subject's gender).

4.3 Example Problems

We can note a few problems that arose with lessons they provide.

One study found that the subjects behaved unexpectedly. The subjects had less problems than was expected. Upon further investigation, it turned out that the student research assistants were breaking up the lessons into subparts to facilitate learning. This is one example that should be avoided to increase validity of research studies by eliminating lesson effects through instruction (e.g., VanLehn, 2007).

4.4 Further Readings

Here is a list of further reading materials concerning this chapter.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.
Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*(98-101).

Cohen originated the current measure of power.

Howell, D. C. (2007). *Statistical methods for psychology* (6th ed.). Belmont, CA: Thomson.

Howell's book provides a useful summary of how to apply power written for those learning statistics. Other introductory statistics book will have similar treatments. They are useful introductions to this process.

5 Running a Research Study

This chapter provides practical information on what to do when you run your experiments. We assume that you have developed your initial experimental design and are now ready to run a pilot study.

5.1 Script

Your research study will likely have a script of how to run the session. If it does not, it should, and it will help you run each subject in a consistent manner. The script will often start with how to setup the apparatus. Before the subject's arrival, the experimenter needs to setup the apparatus and should be ready to welcome the subject. Incorrect or inconsistently applied procedures of the apparatus setup can sometimes cause inconsistency of the study-running processes (e.g., omission of a step). Consequently, the script that appropriately represents required procedures could play an important role to conduct a successful experimental study. Appendix D provides an example script of how to run a study.

5.2 Piloting

As mentioned before, conducting a pilot study based on the script of the research study is important. Piloting can help you determine whether your experimental design will successfully produce scientifically plausible answers to your inquiries. If any revision is necessary, it is far better to find it and correct it before running multiple subjects, particularly when access to subjects is limited. It is, therefore, helpful to think of designing experiments as an iterative process characterized by a cycle of design, testing, and redesign. In addition, you are likely to find that this process works in parallel with other experiments, and may be informed by them (e.g., lessons learned from ongoing related lab work).

We also highly recommend that you use pilot studies to test your written protocols (e.g., instructions for experimenters). The pilot phase provides experimenters the opportunity to test the written protocols with practice participants, and are important for ironing out misunderstandings, discovering problematic features of the testing equipment, and identifying other conditions that might influence the participants. Revisions are a normal part of the process; please do not hesitate to revise your protocols. This will save time later.

It is also useful at this stage to write the method section of your paper. Not only is your memory much fresher but also you can show other researchers your method section and receive suggestions from them. These suggestions can save you a lot of time, in that these reviews essentially constitute another way of piloting the study.

5.3 Dress Code for Experimenters

You should consider the impression you wish to make and will make when running your experiment. This consideration should include how your position, the type of experiment, and the type of participants you are interacting will influence the experiment.

In most cases, we recommend wearing a somewhat formal (or perhaps called business casual) clothing: a dress shirt with dress slacks, when running experiments. This helps you look professional and prepared but not intimidating. Somewhat formal dress helps convey the experiment's importance while not overwhelming the participant. Encouraging your subjects to take the experiment seriously should lead to more distinctive but still generalizable effects.

5.4 Welcome

As the experimenter you are taking on a role similar to that of a host, thus, it is appropriate to welcome participants to the study. Where it is appropriate, you might provide them materials to read if they have to wait, and to answer questions they have before the study begins. It is also very appropriate to confirm their names (for class credit), and to confirm for them that they are in the right place and at the right time. If the experimental protocol permits it, you might also indicate how long the study will take. This helps set the stage for the study itself.

5.5 Missing Subjects

In every study, there are two key parties—the experimenter and the subject or subjects. Inevitably, you will encounter a situation where a participant does not show up despite having an appointment. While participants should notify you in advance if they are going to be absent, keep in mind that missed appointments do happen, and plan around this eventuality. Participants are volunteers (even when you consider compensation). Therefore, it is appropriate to be gracious about their absence. Where possible, we recommend offering to reschedule once. When there are repeated absences, it is often not worth rescheduling.

In some cases, you as an experimenter may need to cancel an experiment. As an experimenter, it is not acceptable to simply not show up for an experiment. When you really have to cancel the experiment for any reason, you should do it in advance. Furthermore as the experimenter, you have the responsibility to cancel the experiment by directly contacting the participants.

5.6 Decorum

Be culturally sensitive and respectful to the participants. Consult with the lead investigator if you have general questions concerning lab etiquette, or specific questions related to the study.

5.7 Debriefing

The APA's ethical principles offer a general outline of debriefing procedures. For many experiments, the lead researcher may provide additional guidance. Investigators should ensure that participants acquire appropriate information about the experiment and the user study—such as the nature, results, and conclusions of the research. If participants are misinformed on any of these points, investigators must take time to correct these misunderstandings. Also, if any procedures are found to harm a participant, the research team must take reasonable steps to alleviate that harm.

The experiment's procedures may cause participants to feel uncomfortable or be alarmed. After the experiment is finished, investigators or experimenters should listen to the participants'

concerns and try to address these problems. Mitchell and Jolley (2007) provide reasonable steps to follow when you need to debrief:

- Correct any misconceptions that participants may have.
- Give a summary of the study without using technical terms and jargon.
- Provide participants an opportunity to ask any questions that they might have.
- Express thankfulness to the participant.

When you have a study that can be perceived as being deceptive or when the study is a double-blind study, you should seek advice about how to debrief the participants. If deception is a procedural component, you will most likely have to explain this to the subjects, and ask that they not discuss the study until the study's completion date. Requesting the participants to refrain from discussing the study will help keep potential subjects from being biased.

To review, double-blind studies prescribe that neither the subject nor the experimenter knows which treatment the subject has received. For example, the amount of caffeine any single participant has ingested in a caffeine study with multiple possible doses. In these cases, you will have to explain the procedures of the study, as well as provide a general rational for double-blind trials. Otherwise, participants may balk at being given a treatment in a sealed envelope, or by a person who is not the experimenter. Furthermore, events such as the Tuskegee and Holmes Prison experiments underscore why procedural transparency is so essential.

5.8 Payments and Wrap-up

At the end of the session, you should be sure to compensate the subject as specified by the disclosure agreement. Compensation can include monetary payment, credit towards a class, or nothing. If you are paying them monetarily, check with your supervisor, as there are nearly always detailed instructions for how to process such payments. In any case, you should make sure that they receive their compensation; you receive any required documentation such as receipts; and that you thank each participant for their assistance. Without them after all, you cannot run the study.

5.9 Simulator Studies

You may find yourself running simulated subjects. User models and simulations are increasingly used, both as standalone objects, but sometimes as part of a study to provide a social context. For example, to model a social situation you might have two intelligent agents act as confederates in a resource allocation game (Nerb, Spada, & Ernst, 1997). These agents provide a known social context in that their behavior is known and can be repeated, either exactly or according to a known set of knowledge.

When you run simulations as subjects, you should keep good notes. There are often differences between the various versions of any simulation, and this should be noted. Simulations will also produce logs, and these logs should be stored as securely and as accurately as subject logs. There may be more of them, so annotating them is very prudent.

5.10 Problems and How to Deal with Them

When you run an experiment, you can encounter unexpected situations in which a participant is exposed to some risk of harm. Investigators must be committed to resolving these problems ethically; recognizing that the well-being of the participants supercedes the value of the study. We recommend consulting your host organization (i.e., Office for Research Protection) in the event that you encounter problems that hinder conducting experiments by affecting either the experimenters or participants. Where these events are adverse enough, you are required to report these events to the IRB board.

5.11 Example Problems

We can note a few problems that we have encountered and some lessons learned.

In one study, we could not run a few subjects because they could not find the room in which we were conducting the experiments. A locked hallway entry door and no escort prevented the participants from finding the room. Ostensibly, the pilot study would have identified this problem; however, only intra-departmental personnel participated in the pilot study. Consequently, the need for an escort was not identified until the first experimental runs. This example highlights the importance of knowing your participant and their needs, as well as the limitations of internal pilot studies.

In another study, a colleague lost valuable data because of a hard drive failure. Spending years gathering data on children, he had not backed up his data. When his hard drive crashed, he was given the choice to spend an extra year rerunning subjects (if support was available), or go to industry. He went to industry where sadly he was very happy!

6 Concluding a Research Session and Study

This section explains practical information about what you should do when you get done with your experiment.

6.1 Data Care, Security, and Privacy

All information and data gathered from an experiment should be considered confidential. If others who are not associated with the experiment have access to either data or personal information, the participants' privacy is violated. Thus, it is the responsibility of lead researchers and experimenters to ensure that all security assurance procedures are promulgated and enforced.

Researchers must safeguard against the inappropriate sharing of sensitive information. Personal information about the participants must not be shared with people not associated with the study. Thus, the data should not be left untended. In most studies, experimental data are kept in locked files or on secure computers. The level of security may vary with the type of data. Anonymous reaction time data, where the only identifying information is a subject ID, is low risk. Personal health records where the subjects might be identified are much more sensitive, and would require more cautious storage, perhaps being used only on a removable disk.

6.2 Data Backup

To protect against data loss, back up all of your data routinely (after running a subject, and every 10 days at minimum when you are doing analyses of the data). If your data is stored in electronic files, store them in a secure hard drive or burn them onto a CD. If you are using paper documents, they can be scanned and stored on a computer file as back up. We suggest that you back up your data after each subject rather than weekly while conducting a study.

6.3 Chance for Insights

Gathering data directly can be tedious, but it can also be very useful. Gathering data gives you a chance to obtain insights about aspects of behavior that are not usually recorded, such as the user's affect, their posture, and their emotional responses to the task.

Obtaining these kinds of insights and the intuition that follows from these experiences is important for everyone, but gathering data is particularly important for young scientists. It gives them a chance to see how previous data has been collected, and how studies work. Reading will not provide you this background or the insights associated with it, rather this knowledge only comes from observing the similarities and differences that arise across multiple subjects in an experiment.

So, be engaged as you run your study and then perform the analysis. These experiences can be a source for later ideas, even if you are doing what appears to be a mundane task. In addition, being vigilant can reduce the number and severity of problems that you and the lead investigator will encounter. Often, these problems may be due to changes in the instrument, or changes due to external events. For example, current events may change word frequencies for a study on

reading. Currently, words such as bank, stocks, and mortgagees are very common, whereas these words were less prevalent three or four years ago.

7 Example Research Studies

We present example studies. In these examples, we show how to plan and prepare to run experiments with human participants. This section can help you obtain practical information for your own study.

7.1 Skill Retention Study

We draw our example from a study investigating the learning and forgetting performance of human participants across a set of procedural tasks—the Office of Naval Research sponsored this work. We present here specific procedures pertaining to the planning and running of the study.

Creation of Study Paradigm: We created a study paradigm to investigate how people forget what they have learned in a laboratory setting. For this study paradigm, we had to create a task that was novel enough to measure learning effects from participants (Kim, 2008). Thus, we chose a free spreadsheet called Dismal (Ritter & Wood, 2005) so that we could minimize any factors that would decrease the data's validity. That is, the task of working with a spreadsheet was familiar to the learners, but the Dismal spreadsheet has not been used by many people. We also needed a tool to unobtrusively record the participant while he or she was performing the task; we chose RUI for this purpose. We then designed the experiment— the study uses a repeated measures design to allow multiple measures of learning and forgetting from each participant.

Here is a list of items that you can use to develop your own study.

- A task
- A task environment
- A tool to record behavior

The First Pilot Test: We tested the study with a couple of pilot subjects. In the first pilot study, we had two male participants. We observed that participants had difficulty in learning the task.

Revising the Task Design: We decided to reduce the cognitive load of the target task so that we could measure learning effects in a restricted time. We reduced the task difficulty so that most of the learning was of the interface and not of the math in the task.

The Second Pilot Test: For the second pilot test, we had a female and a male participant who had no prior knowledge of the task. The task completion times for each were consistent with previous studies. In addition, the participants showed some forgetting over time, which was what we were studying. We concluded that the revised design was satisfactory. In some cases, it might be necessary to iteratively revise the study design further.

Getting IRB Approval: We prepared documents to receive IRB approval. To do this, we needed to provide detailed protocols specifying how we would run the experiments, as well as detailed methods for how we intended to recruit participants.

Start Running Experiments: After getting IRB approval, we started running the main experiment.

In addition to these sequential steps, we would like to share with you some practical information pertaining to the IRB process. When you prepare documents for the IRB, the forms will require you to address the following items. We also include the responses for the learning and retention study.

(a) The benefits of the study:
From your participation, it is expected to obtain data representing how much knowledge and skills can be retained in the memory over time. This research can contribute to design a novel training program.

(b) Any known risks to the participant:
There is no risk to your physical or mental health. During the experiment, you can take a break at any time.

(c) How to achieve the participant's privacy:
Your participation and data are entirely confidential. Personal identification numbers (e.g., PSU ID) will be destroyed after gathering and sorting the experimental data. Without personal identification, the gathered data will be analyzed and used for dissertation and journal publications. The following may review and copy records related to this research: The Office of Human Research Protections in the U.S. Department of Health and Human Services, the Social Science Institutional Review Board, and the PSU Office for Research Protections.

(d) Voluntary participation:
The participation of this study is purely based on volunteerism. You can refuse to answer any questions. At any time, you can stop and decline to continue the experiment. There is no penalty or loss of benefits if you refuse to participate or stop at any time.

(e) Compensation from the experiment:
Participants will receive monetary compensation of $25, $30, or $35 based on your total number of sessions, or extra credits (students registered in IST 331). The experiment consists of 5 to 7 trials ($5 per trial). The compensation will be given as one lump sum after all trials. For the amount of $30 and $35, participants will receive a check issued by Penn State. Others will receive $25 cash. Total research payments within one calendar year that exceed $600 will require the University to annually report these payments to the IRS. This may require you to claim the compensation that you receive for participation in this study as taxable income. For students in IST 331, you will receive 3% added to the total grade in the course. If you do not wish to take part in the research, you may earn the extra credit by completing the following:

- Choose a task to measure your learning and forgetting performance
- Gather your learning and forgetting data for 4 hours of study and test and 1 retention
- Analyze the data to show the total task time per study session and test session

One thing that we want to note here is that investigators should neither implicitly or explicitly force participants to participate in the experiment. As noted above, researchers recruiting students

who are enrolled in their classes must be particularly mindful of how they frame student participation in a study. When investigator use academic credits (extra or otherwise) for compensation, a balanced alternative must be made available.

8 Afterword

There are a many books available about research methods and related statistical analyses. We, however, realized that students usually do not have a chance to learn how to run their own experiments, and that there are no books that teach students practical information about running experiments with human participants.

Students charged with running experiments frequently lack specific domain knowledge in this area. Consequently, young researchers chronically make preventable mistakes. With this book, we hope to assist students as they begin to obtain hands-on knowledge about running experiments. The topics and guidance contained in this book arise from the authors' collective experience in both running experiments and mentoring students.

Further methods of gathering data are being developed. Though these changes will impact the development of future experimental procedures, the gross structures of a study and the aspects we have discussed here are not likely to change.

As you venture into research, you will find new topics that will interest you. We are not able to examine all populations or touch upon measurements and tools that require additional training in this text. Consequently, we are not able to cover in detail the collection of biological specimens, eye-tracking, or fMRI; however with further reading and consultation with colleagues, you will be able to master these skills.

Running studies is often exciting work, and it helps us understand how people think and behave. It offers a chance to improve our understanding in this area. We wish you good luck, bonne chance in finding new scientific results.

Appendix A: Glossary

Independent variable	A variable that is manipulated in the study, either by assignment of materials or assignment of subjects.
Dependent variable	A measurement that is taken during the study, such as reaction time, or percent correct. It depends on other things.
Pilot study	An abbreviated version of the study done to test the procedure and prepare for a larger study.
Power	The power in an experimental study indicates the probability that the test (or experiment) will reject a false null hypothesis. Failure to reject the null hypothesis when the alternative hypothesis is true is referred to as a Type II error. Thus, as the power of a study increases, the chances of a Type II error decrease.
IRB	Internal Review Board. They review study proposals to ensure safety and compliance with US federal regulations.
Informed consent form	
Null hypothesis	The hypothesis that the treatment DOES NOT lead to differences. For example, the null hypothesis might be that two interfaces are equally easy to use.

Appendix B: A Checklist for Setting-up Experiments

As an experimenter or a principal investigator for your project, you need to complete the items below to set up experiments to run.

- ☐ Prepare for the IRB form and submit it to office of research protection
- ☐ Run pilot tests to make sure your experimental design
- ☐ Advertise your experiment to recruit participants (e.g., flyer, a student newspaper)
- ☐ Schedule your participants for an experiment
- ☐ Make sure a lab for the experiment available when you need to run
- ☐ Prepare how to debrief

Appendix C: A Overview of Steps for Running Experiments

As an experimenter or an investigator, you need to consider the items such as those listed below to run your actual experiments.

- ☐ Recruit participants for the experiment
- ☐ Pilot the study
- ☐ Explain detailed information about the experiment (e.g., risks, benefits, and the purpose of the experiment)
- ☐ Let participants know they can stop participation and performance at any time
- ☐ Protect the participant's confidentiality
- ☐
- ☐ Be ready to address any harms or risks from the experiment by talking with the lead investigator
- ☐
- ☐ Make sure ethical codes by APA are being considered
- ☐ Archive the data
- ☐ Analyse data
- ☐ Report data

Appendix D: Example Script to Run an Experiment

This is one page script that every experiment should read and follow.

Experimenter's Guide

This is an example script for an experiment. Every experimenter should follow the procedures to run a user study about skill retention.

(1) Check your dress code
(2) Before your participants are coming in, you need to set up a set of the experiment apparatus.
   a) Start RUI in the Terminal Window. (see details ..)
   b) Start the Emacs text editor.
   c) Prepare disposable materials, handouts, such as informed consent form
(3) Welcome your participants
(4) Put a sign on the door indicating that you are running subjects when the experiment starts
(5) Give the IRB form and have them read it
(6) If they consent to it, start the experiment
(7) Briefly explain what they are going to do
(8) Give them the study booklet.
   a) Participants can use 30 min maximum to study the booklet.
(9) While participants are reading the booklet, you can answer their questions about the task.
(10) Turn on the monitor that is located in the experimental room, so that you can monitor the participant outside the room.
(11) When the experiment is finished, give an explanation about the payments or extra credit. Also, if there are any additional schedules for later measures, remind them.
(12) Take down the sign on the door when the experiment is done
(13) Copy the data to external hard drive
(14) Shut down apparatus
(15) Make supplies for next subject

Using RUI

RUI (Recording User Input) will be used to log keystrokes and mouse actions of the participant. RUI requires Mac OS X 10.3 (Panther) or later versions. It has been tested up to Mac OS X 10.4.3. (Tiger). In order for RUI to record user inputs, "Enable access for assistive devices" must be enabled in the Universal Access preference pane.

(1) Launch Terminal
(2) In Terminal, type the below information:
   ./rui –s "Subject Name" –r ~/Desktop/ruioutput.txt
(3) You will get this message:
   rui: standing by – press ctrl+r to start recording…
(4) Press "CTRL+r"
(5) To stop recording, press "CTRL+s"

Note:
If you see the message of "-bash: ./rui: Permission denied" in the Terminal window, you need to type "chmod a+x rui" while you are in the RUI directory.

Measuring Learning & Forgetting

Emacs is started by the experimenter for every session. The participants will start and stop RUI to record their performance. The experimenter needs to ensure that the participants cannot do mental rehearsal during the retention period.

Appendix E: Safety of Experiments

Some common safety concerns:
for cog psy, there are none

for interesting things, see your irb
for stress, see your irb
for taking samples from humans, see your IRB

Appendix F: Example Consent Form

Here is an example of an informed consent form that you can refer to when you need to generate one for your experiment.

ORP USE ONLY: IRB#21640 Doc. #1
The Pennsylvania State University
Office for Research Protections
Approval Date: 09/09/2008 – J. Mathieu
Expiration Date: 09/04/2009 – J. Mathieu
Biomedical Institutional Review Board

Informed Consent Form for Biomedical Research
The Pennsylvania State University

Title: Investigating a Forgetting Phenomenon of Knowledge and Skills

Principal Investigator: Dr. Frank E. Ritter
316G IST Bldg, University Park, PA 16802
(814) 865-4453 frank.ritter@psu.edu

Other Investigators:

Dr. Jong Wook Kim
316E IST Building
University Park, PA 16802
(814) 865-6166; jongkim@psu.edu

Dr. Richard J. Koubek
310 Leonhard Building
University Park, PA 16802
(814) 865-7601 rkoubek@psu.edu

1. **Purpose & Description:** The purpose of the study is to investigate how much knowledge and skills are forgotten and retained in human memory after a series of learning sessions. Human performance caused by forgetting will be quantitatively measured. If you decide to take part in this experiment, please follow the experimenter's instruction.

   The experiment is held at 319 (Applied Cognitive Science Lab.) or 205 (a computer lab) IST building. During the experiment, the timing of keystrokes and mouse movements will be recorded.

   A group of participants (80 participants) selected by chance will wear an eye-tracker to measure eye movements during the task, if you consent to wear the device. You can always refuse to use it. The eye-tracker is a device to measure eye positions and eye movements. The eye-tracker is attached to a hat, so you just can wear the hat for the experiment. The device is examined for its safety. You may be asked to talk aloud while doing the task.

2. **Procedures to be followed:**
   a. You will be asked to study an instruction booklet to learn a spreadsheet task (e.g., data normalization). Each study session will be 30 minutes maximum. For four days in a row, you will learn how to do the spreadsheet task.
   b. Then, you will be asked to perform the given spreadsheet tasks on a computer (duration: approximately 15 minutes).
   c. With a retention interval of 6-, 9-, 12-, 18-, 30-, or 60-day, after completing the second step, you will be asked to return to do the same spreadsheet task (duration: approximately 15 min/trial)

3. **Voluntary Participation:** The participation of this study is purely based on volunteerism. You can refuse to answer any questions. At any time, you can stop and decline the experiment. There is no penalty or loss of benefits if you refuse to participate or stop at any time.

4. **Right to Ask Questions:** You can ask questions about this research. Please contact Jong Kim at jongkim@psu.edu or 814-865-6166 with questions, complaints, concerns, or if you feel you have been harmed by this research. In addition, if you have questions about your rights as a research participant, contact the Pennsylvania State University's Office for Research Protections at (814) 865-1775.

5. **Discomforts & Risks:** There is no risk to your physical or mental health. You may experience eye fatigue because you are interacting with a computer monitor. During the experiment, you can take a break at any time.

6. **Benefits:** From your participation, it is expected to obtain data representing how much knowledge and skills can be retained in the memory over time. This research can make a contribution to design a novel training program.

7. **Compensation:** Participants will receive monetary compensation of $25, $30, or $35 in terms of your total trials, or extra credits (students registered to IST 331). The experiment consists of 5 to 7 trials ($5 per trial). The compensation will be given as one lump sum after all trials. For the amount of $30 and $35, participants will receive a check issued by Penn State. Others will receive a cash of $25. Total research payments within one calendar year that exceed $600 will require the University to annually report these payments to the IRS. This may require you to claim the compensation that you receive for participation in this study as taxable income.

8. **Confidentiality:** Your participation and data are entirely confidential. Personal identification numbers (e.g., PSU ID) will be destroyed after gathering and sorting the experimental data. Without personal identification, the gathered data will be analyzed and used for dissertation and journal publications. The following may review and copy records related to this research: The Office of Human Research Protections in the U.S. Department of Health and Human Services, the Social Science Institutional Review Board and the PSU Office for Research Protections.

You must be 18 years of age or older to take part in this research study. If you agree to take part in this research study and the information outlined above, please sign your name and indicate the date below.

You will be given a copy of this signed and dated consent for your records.

_____ _____

Participant Signature Date

_____ _____

Person Obtaining Consent (Principal Investigator) Date

Appendix G: Example Debrief Form

HRI Debriefing Form

Thank you for participating in our human-robot interface testing study.

From your participation we will learn how people use interfaces in general and Human-Robot interfaces in particular. These interfaces are similar to those used to interfaces used to work in hazardous areas including those used in rescue work at the World Trade Center. By participating, you have been able to see and use a new technology. The results can lead to improved interfaces for robots that replace humans in hazardous conditions.

You may also find the Robot project overview page useful and interesting.

If you have any questions, please feel free to ask the experimenter. You can also direct questions to Dr. Frank Ritter, (frank.ritter@psu.edu, 865-4453).

Appendix H: Example IRB Application

Your Internal Review Board will have its own review forms. These forms are based on each IRB's institutional history, and the types of studies and typical problems (and atypical problems) that they have had to consider over time. Thus, the form we include here can only be seen as an example form. We include it to provide you with an example of the types of questions and more importantly the types of answers characteristic of the IRB process. You are responsible for the answers, but it may be useful to see examples to see how long they are, and how detailed they need to be.

Following is a form used in one of our recent studies.

PENNSTATE

Office for Research Protections
201 Kern Building
University Park, PA 16802
814-865-1775
Fax: 814-863-8699
ORProtections@psu.edu

APPLICATION FOR THE USE OF HUMAN PARTICIPANTS
EXPEDITED & FULL REVIEWS

OFFICE USE ONLY
IRB NO._____

Form Instructions:
- To complete the form, press TAB or SHIFT TAB between boxes and enter an 'X' or text. For assistance, contact the Office for Research Protections.
- This application will ask general questions about your study. Depending on your response, additional appendices may need to be completed in order to provide more detailed information. For example, if you indicate that your study involves prisoners, **Appendix 4** will also need to be completed and submitted.
- Submit recruitment materials, informed consent forms, and all other materials as attachments to the application. **Do NOT** include within the application.
- Handwritten applications will NOT be accepted.

Project Title: **Gathering Data From Computer Interface Users to Test Cognitive Models**

Principal Investigator: **Frank Ritter, PhD, C. Psychol.**	PSU User ID (e.g., abc123): **fer2**
University Status (Faculty, Staff, Student, etc.): **Faculty**	Telephone Number: **+1 (814) 865-4453**
Email Address: **frank.ritter@psu.edu**	Dept: None
College: **College of IST**	Campus: **University Park**
Mailing Address: **316G Building IST**	

Faculty Advisor, if PI is a student:	PSU User ID (e.g., abc123):
Email Address:	Telephone Number:
Dept:	College:
Mailing Address:	Campus:

Is there anyone you wish to include on correspondence related to this study (e.g., a study coordinator, etc.)?

Name:	PSU User ID (e.g., abc123):
University Status (Faculty, Staff, Student, etc.):	Telephone Number:
Email Address:	Dept:
College:	Campus:
Mailing Address:	Role in this study: **Choose one of the following**

A. Funding:

1. Is this research study internally or externally funded?
   - ☐ Yes Answer Questions 2 – 4
   - ☒ No Skip to Question 6
   - ☐ Pending Answer Questions 2 – 5

2. Provide the name and mailing address of internal and external sources of funding. Provide a copy of your grant proposal with the application. If a copy of the grant proposal is not included, explain.

3. Is the sponsor providing the drug, device, etc. free of charge? ☐ Yes ☐ No ☐ N/A

4. Has the sponsor agreed to pay for direct costs of treating injuries? ☐ Yes ☐ No

5. If funding is not awarded, will the research still be conducted? ☐ Yes ☐ No ☐ N/A

B. Conflict of Interest:

6. Do any of the investigator(s), key personnel, and/or their spouses or dependent children have a conflict of interest (COI), as defined by PSU Policy RA20, "Individual Conflict of Interest," associated with this research?
   - ☐ Yes Complete & Submit **Appendix 1, Section A**
   - ☒ No

7. Does PSU have an ownership or royalty interest in any intellectual property related to this study?
   - ☐ Yes Complete & Submit **Appendix 1, Section B**
   - ☒ No

8. Are there are other significant conflicts that could possibly affect or be perceived to affect this study?
   - ☐ Yes Complete & Submit **Appendix 1, Section C**
   - ☒ No

C. Class Projects:

9. Is this a class project?
   - ☐ Yes Provide the following information:
     - Instructor's Name:
       Course Title and Number:
       Semester course is being offered:
   - ☒ No

D. Review Level:

10. What level of review do you expect this research to need?
    - ☒ Expedited Review Answer Question 11
    - ☐ Full Review Skip to Question 12

11. Expedited Research Categories: Read the following categories and choose one or more that apply to your research. Your research must fit in at least one category and be no more than minimal risk in order to be considered for an expedited review.

☐ **Category 1:** Clinical studies of drugs and medical devices only when condition (a) **OR** (b) is met.

(a) Research on drugs for which an investigational new drug application (21 CFR 312) is not required. *(Research on marketed drugs that significantly increases the risks or decreases the acceptability of the risks associated with the use of the product is not eligible for expedited review.)*

(b) Research on medical devices for which (i) an investigational device exemption application (21 CFR 812) is not required; or (ii) the medical device is cleared/approved for marketing and the medical device is being used in accordance with its cleared/approved labeling.

☐ **Category 2:** Collection of blood samples by finger stick, heel stick, ear stick or venipuncture as follows:

(a) From healthy, non-pregnant adults who weigh at least 100 pounds. For these participants, the amounts drawn may not exceed 550 ml in an 8 week period and collection may not occur more frequently than 2 times per week; **OR**

(b) From other adults and children, considering the age, weight, and health of the participants, the collection procedure, the amount of blood to be collected, and the frequency with which it will be collected. For these participants, the amount drawn may not exceed the lesser of 50 ml or 3 ml per kg in an 8 week period and collection may not occur more frequently than 2 times per week.

☐ **Category 3:** Prospective collection of biological specimens for research purposes by non-invasive means. Examples include:

- Hair and nail clippings in a non-disfiguring manner;
- Deciduous teeth at time of exfoliation or if routine patient care indicates a need for extraction;
- Permanent teeth if routine patient care indicates a need for extraction;
- Excreta and external secretions (including sweat);
- Uncannulated saliva collected either in an unstimulated fashion or stimulated by chewing gumbase or wax or by applying a dilute citric solution to the tongue;
- Placenta removal at delivery;
- Amniotic fluid obtained at the time of rupture of the membrane prior to or during labor;
- Supra- and subgingival dental plaque and calculus, provided the collection procedure is not more invasive than routine prophylactic scaling of the teeth and the process is accomplished in accordance with accepted prophylactic techniques;
- Mucosal and skin cells collected by buccal scraping or swab, skin swab, or mouth washings;
- Sputum collected after saline mist nebulization

☐ **Category 4:** Collection of data through non-invasive procedures (not involving general anesthesia or sedation) routinely employed in clinical practice, excluding procedures involving x-rays or microwaves. Where medical devices are employed, they must be cleared/approved for marketing. Studies intended to evaluate the safety and effectiveness of the medical device are not generally eligible for expedited review, including studies of cleared medical devices for new indications. Examples include:

- Physical sensors that are applied either to the surface of the body or at a distance and do not involve input of significant amounts of energy into the participant or an invasion of the participant's privacy;
- Weighing or testing sensory acuity;
- Magnetic resonance imaging;
- Electrocardiography, electroencephalography, thermography, detection of naturally occurring radioactivity, electroretinography, ultrasound, diagnostic infrared imaging, doppler blood flow, and echocardiography;
- Moderate exercise, muscular strength testing, body composition assessment, and flexibility testing where appropriate given the age, weight, and health of the individual.

☐ **Category 5:** Research involving materials (data, documents, records, or specimens) that have been collected or will be collected solely for non-research purposes (such as medical treatment or diagnosis).

☒ **Category 6:** Collection of data from voice, video, digital, image recordings made for research purposes.

☐ **Category 7:** Research on individual or group characteristics or behavior (including but not limited to research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

E. Research Personnel:

NOTE:

- The Principal investigator is responsible for ensuring that all individuals conducting procedures described in this application are trained adequately prior to involving human participants.
- All personnel listed on this application who (1) are responsible for the design/conduct of the study, (2) will have access to the human participants (i.e., will consent participants, conduct the study), or (3) will have access to identifying AND confidential information must successfully complete the IRB's Training on the Protection of Human Participants or provide verification of training from their home institution. PSU's training may be located at http://www.research.psu.edu/orp/education/modules/irb/index.asp. **Approval will NOT be granted until all individuals have successfully completed the training.** Verification of training does NOT need to be sent in if the individual completed the Penn State's training.
- As personnel change, you must submit a *Modification Request Form – Expedited & Full Review* to add or remove personnel.

☐ **Category 2:** Collection of blood samples by finger stick, heel stick, ear stick or venipuncture as follows:

(a) From healthy, non-pregnant adults who weigh at least 100 pounds. For these participants, the amounts drawn may not exceed 550 ml in an 8 week period and collection may not occur more frequently than 2 times per week; **OR**

(b) From other adults and children, considering the age, weight, and health of the participants, the collection procedure, the amount of blood to be collected, and the frequency with which it will be collected. For these participants, the amount drawn may not exceed the lesser of 50 ml or 3 ml per kg in an 8 week period and collection may not occur more frequently than 2 times per week.

☐ **Category 3:** Prospective collection of biological specimens for research purposes by non-invasive means. Examples include:

- Hair and nail clippings in a non-disfiguring manner;
- Deciduous teeth at time of exfoliation or if routine patient care indicates a need for extraction;
- Permanent teeth if routine patient care indicates a need for extraction;
- Excreta and external secretions (including sweat);
- Uncannulated saliva collected either in an unstimulated fashion or stimulated by chewing gumbase or wax or by applying a dilute citric solution to the tongue;
- Placenta removal at delivery;
- Amniotic fluid obtained at the time of rupture of the membrane prior to or during labor;
- Supra- and subgingival dental plaque and calculus, provided the collection procedure is not more invasive than routine prophylactic scaling of the teeth and the process is accomplished in accordance with accepted prophylactic techniques;
- Mucosal and skin cells collected by buccal scraping or swab, skin swab, or mouth washings;
- Sputum collected after saline mist nebulization

☐ **Category 4:** Collection of data through non-invasive procedures (not involving general anesthesia or sedation) routinely employed in clinical practice, excluding procedures involving x-rays or microwaves. Where medical devices are employed, they must be cleared/approved for marketing. Studies intended to evaluate the safety and effectiveness of the medical device are not generally eligible for expedited review, including studies of cleared medical devices for new indications. Examples include:

- Physical sensors that are applied either to the surface of the body or at a distance and do not involve input of significant amounts of energy into the participant or an invasion of the participant's privacy;
- Weighing or testing sensory acuity;
- Magnetic resonance imaging;
- Electrocardiography, electroencephalography, thermography, detection of naturally occurring radioactivity, electroretinography, ultrasound, diagnostic infrared imaging, doppler blood flow, and echocardiography;
- Moderate exercise, muscular strength testing, body composition assessment, and flexibility testing where appropriate given the age, weight, and health of the individual.

☐ **Category 5:** Research involving materials (data, documents, records, or specimens) that have been collected or will be collected solely for non-research purposes (such as medical treatment or diagnosis).

☒ **Category 6:** Collection of data from voice, video, digital, image recordings made for research purposes.

☐ **Category 7:** Research on individual or group characteristics or behavior (including but not limited to research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

E. Research Personnel:

NOTE:

- The Principal investigator is responsible for ensuring that all individuals conducting procedures described in this application are trained adequately prior to involving human participants.
- All personnel listed on this application who (1) are responsible for the design/conduct of the study, (2) will have access to the human participants (i.e., will consent participants, conduct the study), or (3) will have access to identifying AND confidential information must successfully complete the IRB's Training on the Protection of Human Participants or provide verification of training from their home institution. PSU's training may be located at http://www.research.psu.edu/orp/education/modules/irb/index.asp. **Approval will NOT be granted until all individuals have successfully completed the training.** Verification of training does NOT need to be sent in if the individual completed the Penn State's training.
- As personnel change, you must submit a *Modification Request Form – Expedited & Full Review* to add or remove personnel.

12. Provide the name of the other individual(s) assisting with this study who (1) will be responsible for the design/conduct of the study, (2) have access to the human participants (i.e., will consent participants, conduct the study), or (3) have access to identifying AND confidential information. If the individual does not have a PSU Access User ID, please provide some other form of contact information. If additional space is needed, attach a separate sheet containing the same information.

Name	Email Address	PSU User ID (e.g., abc 123)	Mailing Address	Role in this Study
Maik Friedrich	mFriedrich@ist.psu.edu	muf16	College of IST, Penn State, University Park, PA 16802	Research Assistant
				Choose one of the following
				Choose one of the following
				Choose one of the following
				Choose one of the following
				Choose one of the following
				Choose one of the following

13. Identify (1) the procedures/techniques each person (including advisors) listed in Question 12 and on the first page of the application will perform and (2) describe their level of research experience.

(1) Each person will recruit subjects for a cognitive psychology study, to run the subjects in a cognitive psychology study, and to store the data.
(2) Ritter has taught these research methods and run studies in this area since 1990. Friedrich has passed the PSU IRB test and has been instructed by Ritter, and will continue to be trained.

14. Explain how the persons assisting with this research are kept adequately informed about the study and their research-related duties and functions.

The persons will be instructed in how to perform cognitive psychology studies, and their performance will be monitored by Ritter. This will include discussions following running the first five subjects.

F. Purpose & Procedures:

15. Provide a detailed description of the research that includes (1) the background, (2) aims/objectives [hypothesis], and (3) a description of how the research will be conducted [methodology – what participants will be asked to do].

This research is designed to look at how people problem solve, and how they learn. The task used will be a previously used simple interface problem solving task. (2) The hypothesis are related to how fast the learning will occur, and are an exploration to discover what strategies users use. (3) The participants will be asked to read instructional materials, and then to solve faults in a simple device. Their behavior will be recorded while they do this.

16. How long will participants be involved in this research study? Include the number of sessions and the duration of each session.

1 sessions, 30 to 50 minutes per session.

17. Where will this research study take place? **Choose all that apply.**

- ☒ University Park Specify the building and room number. If not yet known, indicate such. **319b IST Building**
- ☐ GCRC at University Park
- ☐ Other PSU Campus Location Specify the campus, building and room number. If not yet known, indicate such.
- ☐ Hershey Medical Center Specify the building and room number. If not yet known, indicate such.
- ☐ GCRC at the Hershey Medical Center
- ☐ Mt. Nittany Medical Center
- ☐ Other Site(s) Explain:

NOTE: For other sites such as schools, doctor offices, businesses, etc., the IRB requires that research conducted at these sites be approved by an individual in a decision making position at the site. Documented approval (i.e., a letter of agreement) is required.

18. Is this a multi-center study outside of PSU?
☐ Yes Answer Question 19
☒ No Skip to Question 22

19. Is any Penn State investigator on this application the lead investigator (project director) of this multi-center study?
☐ Yes Answer Questions 20 – 21
☐ No Skip to Question 22

20. Provide the name and location of all other centers. Copies of IRB approval letters from each site will be required with the supporting documentation for this application.

21. Describe the plan for the management and communication of multi-site information that may be relevant to the protection of participants (e.g., unanticipated problems, adverse events, interim analyses, modifications).

22. How will the data be analyzed?
The data will be summarized across problem series; the data will be compared to existing predictions of a computational model; the pattern of the fit to the predictions will be used to find and define new strategies.

23. List criteria for inclusion of participants.
Willingness to participate in study. Over 18 years old.

24. List criteria for exclusion of participants.
None.

G. Participants:

25. Maximum number of participants/samples/charts to be enrolled at this institution (Enter one number – not a range): **20**

26. Was a statistical/power analysis conducted to determine the adequate sample? ☐ Yes ☐ No **NA This study does not use inferential statistics.**

27. Does this research exclude any particular:

Gender Identity	☐ Yes	☒ No	If Yes, please explain.
Racial/ethnic groups	☐ Yes	☒ No	If Yes, please explain.
Sexual Orientation	☐ Yes	☒ No	If Yes, please explain.

28. Age range – **Choose all that apply.**

☐ Less than 1 year	☐ 7 – 12 years	☒ 18 – 25 years	☒ 40 – 65 years
☐ 1 – 6 years	☐ 13 – 17 years	☒ 26 – 40 years	☒ 65+ years

29. **Choose all categories of participants** who will be involved in this research study.
☒ Healthy volunteers
☒ Penn State students
☐ Subject Pool Students – Indicate the subject pool: ☐ CAS 100A ☐ Psychology – UP ☐ Psychology – Behrend
Will all participants involved in this study be from the subject pool? ☐ Yes ☐ No
☐ Children – Individuals under the age of 18 Complete & Submit **Appendix 2**
☐ International Research – participants live outside of the U.S. Complete & Submit **Appendix 3**
☐ Prisoners Complete & Submit **Appendix 4**
☐ Pregnant Women
☒ Women of reproductive potential at the time of this research – Choose one of the following:
☒ The research poses no added risk associated with pregnancy and/or lactation
☐ Precautions against pregnancy and/or lactation, and pregnancy tests are addressed in the research proposal and consent form

☐ Patients — Complete & Submit **Appendix 5**
☐ Individuals with a decisional impairment who are targeted for this study (e.g., research on Alzheimer's enrolling only individuals with Alzheimer's) — Complete & Submit **Appendix 6**
☐ Individuals with a decision impairment who are NOT targeted for this study (e.g., decisionally compromised person eligible for a study on a new treatment for breast cancer) — Complete & Submit **Appendix 6**
☐ Institutionalized individuals (e.g., patients in state hospitals or nursing homes) — Complete & Submit **Appendix 7**
☐ Fetus, embryo, fetal material in vitro fertilization
☐ None of the above categories will be used in this research

30. Will participants be currently enrolled in a course/class of any personnel listed on this application?
☐ Yes Describe the measures taken to avoid coercion & undue influence:
☒ No

31. Will participants be employees of any personnel listed on this application?
☐ Yes Describe the measures taken to avoid coercion & undue influence:
☒ No

32. Could some or all participants be vulnerable to coercion or undue influence due to special circumstances? Do not include children, decisionally impaired persons, and prisoners in your answer.
☐ Yes Describe the measures taken to protect these individuals:
☒ No

H. Recruitment:

33. Indicate the types of recruitment that will be done for this research & **attach copies of the materials**. Choose all that apply:
☐ Newspaper/magazine ads
☐ Radio/TV ads
☐ Letters/Emails to potential participants
Explain how potential participants contact information was obtained:
☐ Letters/Emails to healthcare professionals for recruitment purposes
Which healthcare groups will receive these letters?
☒ Flyers/posters – Where will the items be displayed/distributed? **In the college of IST**
☐ Brochures – Where will the items be displayed/distributed?
☐ Web sites – List the sites the recruitment materials will be posted:
☐ Email via Listserv – Has permission been obtained from the listserv administrator? ☐ Yes ☐ No
☐ Script – Verbal (i.e., telephone, face-to-face, classroom)
☐ Subject Pool Indicate which subject pool will be used:
☐ CAS 100A ☐ Psychology – UP ☐ Psychology – Behrend
Note: If you are not a member of the subject pool's department, a permission letter will be needed.
☐ Other Explain:

34. Who will approach and/or respond to potential participants?
Ritter and Friedrich

35. Before potential participants sign a consent form, are there any screening questions that will be asked to determine whether an individual is appropriate for the study?
☐ Yes Answer Question 36
☒ No Skip to Question 37

36. During screening questions, will identifiable information about these individuals be recorded?
☐ Yes Complete & Submit **Appendix 8**
☐ No

NOTE: Please attach, as appropriate, a procedure and script for the screening questions. Also, attach a copy of the screening question data collection sheet.

37. Will investigators access medical charts and/or hospital/clinic databases for recruitment purposes?
- ☐ Yes Answer Question 38
- ☒ No Skip to Question 39

38. Has a waiver of authorization to access protected health information been requested?
- ☐ Yes
- ☐ No Explain why a waiver of authorization has NOT been requested:

39. Will physicians/clinicians provide identifiable, patient information (e.g., name, telephone number, address) to investigators for recruitment purposes?
- ☐ Yes Provide a copy of the written authorization release form for review.
- ☒ No

I. Consent:

40. When and where will participants be approached to obtain informed consent/assent [include the timing of obtaining consent in the response]? If participants could be non-English speaking, illiterate or have other special circumstances, describe. **Attach a copy of the informed consent/assent form(s).**

Participants will be approached to obtain informed consent when they come to participate in the study.

41. Who will be responsible for obtaining informed consent/assent from participants?

The experimentor running the session. Ritter or the RA.

42. Do the people listed in Question 41 above speak the same language as the participants?
- ☒ Yes
- ☐ No Explain how consent will be obtained.

43. What type of consent will be obtained? **Choose all that apply.**
- ☒ Signed consent – participant will sign consent form
- ☐ Implied consent – participant will not sign consent form (e.g., mail survey, email, on-line survey)
  Complete & Submit **Appendix 9, Section A**
- ☐ Verbal consent – participant gives consent verbally (e.g., in-person interview, telephone interview)
  Complete & Submit **Appendix 9, Section A**
- ☐ Passive/Opt Out consent – participant only required to act if they do not want to participate
  Complete & Submit **Appendix 9, Section B**
- ☐ Complete waiver of informed consent
  Complete & Submit **Appendix 9, Section B**
- ☐ Other Describe:

44. If multiple groups of participants are being utilized (i.e., teachers, parents, children, people over 18), who will and will not sign the assent/consent form? Specify for each group of participants.

45. Participants are to receive a copy of the informed consent form with the approval box/statement on it. Describe how participants will receive a copy of the informed consent form to keep for their records.

They will sign two copies, and one copy will be handed to them.

J. Payment for Participation:

46. Indicate the type and amount of payment for participation that will be offered. **Choose all that apply.**

☒ Money	Amount:	**$7**	Skip to Question 48
☐ Gift Certificate	Amount:		Skip to Question 48
☐ Extra/Class Credit (e.g., 5 points, 1% of final grade)	Amount:		Skip to Question **47**
☐ Drawing	Explain:		Skip to Question 48

☐ Other (e.g., merchandise) Explain: Skip to Question 48
☐ Compensation will **NOT** be offered Skip to Question **49**

47. An alternative, equal in time and effort, must be offered in place of participating in the research. Describe the alternative available for earning the extra/class credit. The description should include the length of time it will take to complete the alternative as well as how undue influence will be prevented.

48. Will compensation be pro-rated? NOTE: Pro-rating is required for FDA-regulated studies.
☐ Yes Explain how payment will be pro-rated:
☒ No

K. Data Collection Measures/Instruments:

49. **Choose any of the following** data collection measures/instruments that will be used in this study. **Attach a copy of all instruments/measures, interview and focus group topics/questions to the application.**
☐ Biological Specimens – blood, urine & other human derived samples
☐ Biomedical Devices – EEG, EKG, MRI
☐ Diaries/Journals completed by the participants
☐ Focus Groups
☐ Individual Interviews
☐ Knowledge/Cognitive Tests
☐ Observations
☐ Physical Testing Measures – Height, Weight, Body Mass Index, Blood Pressure
☐ Questionnaires/Surveys – Mail, Internet, Telephone, Email, Paper/Pencil
☒ Other Explain: Mouse moves, keystrokes, and video recordings.

50. Will participants be assigned to groups?
☐ Yes Answer Questions 51 – 52
☒ No Skip to Question 54

51. Will a control group(s) be used?
☐ Yes **Choose one of the following:**
☐ Placebo control
☐ Standard therapy control
☐ Other control method Explain:
☐ No

52. Is the research a blinded (masked) study?
☐ Yes Answer Question 53
☐ No Skip to Question 54

53. Is emergency unblinding permitted?
☐ Yes
☐ No Explain why emergency unblinding is NOT permitted:

L. Recordings – Audio, Video, Photographs

54. Will any type of recordings (audio or video) or photographs be made during this study?
☒ Yes Complete & Submit **Appendix 10**
☐ No

M. Computer/Internet

55. Will any participant interaction in this study be conducted on the Internet or via email (e.g., on-line surveys, observations of chat rooms or blogs, on-line interviews)?
☐ Yes Complete & Submit **Appendix 11, Section A**
☒ No

56. Will a commercial server (i.e., SurveyMonkey, Psych Data, Zoomerang) be used to collect data or for data storage?
☐ Yes Complete & Submit **Appendix 11, Section B**
☒ No

N. Discomforts and Risks

57. List all of the potential discomforts and risks (physical, psychological, legal, social or financial) and describe the likelihood or seriousness of the discomforts/risk. If there are no discomforts/risks, state such.
No additional discomforts or risks for participants beyond daily life.

58. Describe how risks will be minimized and/or how participants will be protected against potential risks throughout the study.
There are no additional risks to participants.

59. Does this research involve greater than minimal risk to the participants?
☐ Yes Answer Questions 60 – 61 **Study must be reviewed by the Full IRB at a convened meeting.**
☒ No Skip to Question 62

60. Will medical or psychological care be available for participants who may require it as a result of the study?
☐ Yes Identify the source of medical or psychological care available – include address & telephone number:
☒ No Explain why medical or psychological care will NOT be available: **It is difficult to choose what possible care could be required.**

61. Does the research protocol have a plan for routine analysis or monitoring of the data and safety of this research study?
☐ Yes Complete & Submit **Appendix 17**
☒ No For studies involving greater than minimal risk, a plan will need to be developed for review and approval at the convened IRB meeting.

O. Benefits

62. What are the potential benefits to the individual participants? If none, state such. PLEASE NOTE: Payment for participation cannot be considered a benefit.
There are two benefits put forward. Participants will get to see how a study is performed, and the data that is gathered may lead to improved instructional material for them and for society.

63. What are the potential benefits to society? If none, state such.
The results from the study can improve our understanding of how learning and problem solving occurs. This has implications for training and education.

64. Explain how the benefits outweigh the risks.
There are minimal risks, and the benefits are that this study can lead to improved training and learning paradigms.

P. Reporting

65. Is it possible investigators will discover a participant's previously unknown condition (e.g., disease, suicidal thoughts, wrong paternity) as a result of study procedures?
☐ Yes Explain how and when such a discovery will be handled:
☒ No

66. Is it possible investigators will discover a participant is engaging in illegal activities (e.g., drug use, domestic violence, child abuse/neglect, underage drinking) as a result of study procedures?
☐ Yes Explain how and when such a discovery will be handled:
☒ No

Q. Deception

67. Does this study involve giving false or misleading information to participants or withholding information from them such that their "informed" consent is in question?

☐ Yes Complete & Submit **Appendix 12**
☒ No

R. Confidentiality and Privacy

68. Describe the provisions made to maintain confidentiality of the data. **Choose all that apply.**
☐ Password protected computer files ☒ Locked offices
☐ Locked file cabinets ☐ Other Explain:
☒ Identification code (i.e., code numbers, pseudonyms) – data will NOT be associated w/personal identifiers

69. Describe the provisions made to protect participants' privacy interests.
No identifying information will be kept with the keystroke logs or with the videos.

70. Who will have access to the data?
Researchers approved to work on this study.

71. Will identifiers be disclosed to a sponsor or collaborators at another institution?
☐ Yes List the identifiers that will be disclosed and explain why this is necessary:
☒ No

72. Will a list containing a code (i.e., code numbers, pseudonyms) and participants' identity be used in this study?
☒ Yes Answer Questions 73 – 75
☐ No Skip to Question 76

73. Where will the list linking the code to participants' identity be stored and how will the list be secured?
List will be kept on a printed sheet in the PI's office.

74. Who will have access to the list linking the code to participants' identity?
the RA when running the study and PI after the study has been run.

75. Will the list linking the code to participants' identity be destroyed?
☒ Yes When will the list be destroyed? **5 years after running the study or after the last publication, whichever comes last.**
☐ No

76. What will happen to the research records when the research has been completed? **Choose only one.**
☐ Stored indefinitely with identifiers removed
☒ Stored indefinitely with identifiers attached
List the identifiers that will be attached to the data: **subject ID**
Explain why the data must be stored indefinitely with identifiers: **once an ID has been assigned, analyses use that ID**
☐ Stored for length of time required by federal regulations/funding source & then destroyed (minimum of 3 years)
☐ Destroyed after a number of years (minimum of 3 years) Specify the number of years:
☐ Destroyed when notified by sponsor
☐ Other Explain:

77. Could the information being collected for this study have adverse consequences for participants or be damaging to their financial standing, employability, insurability or reputation?
☐ Yes Indicate the type of information being collected:
☐ Substance abuse or other illegal risk behaviors
☐ Determination of HIV status for the research
☐ Genetic information about inheritable diseases
☐ Other Explain:
☒ No

78. Will a "Certificate of Confidentiality" be obtained from the federal government?
☐ Yes Indicate who will obtain the Certificate of Confidentiality

☐ Sponsor
☐ Principal Investigator
☐ Other Explain:
☒ No

S. Health Insurance Portability & Accountability Act (HIPAA) – Use of protected health information

79. Will participant's protected health information (PHI) be obtained for this study?
☐ Yes Complete & Submit **Appendix 13**
☒ No

T. Drugs, Medical Devices, and Other Substances

80. Does this research study involve drugs or biologics?
☐ Yes Complete & Submit **Appendix 14, Section A**
☒ No

81. Does this research study involve a device?
☐ Yes Go to Question 82
☒ No Skip to Question 83

82. Does the device meet the FDA's definition of a medical device?
☐ Yes Complete & Submit **Appendix 14, Section C**
☐ No Go to Question 83

FDA's Definition of a Medical Device: If a product is labeled, promoted or used in a manner that meets the following definition in section 201(h) of the Federal Food Drug and Cosmetic (FD&C) Act it will be regulated by the Food and Drug Administration (FDA) as a medical device and is subject to pre-marketing and post-marketing regulatory controls. A device is:

- "an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including a component part, or accessory which is:
  - Recognized in the official National Formulary, or the United States Pharmacopoeia, or any supplement to them,
  - Intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or
  - Intended to affect the structure or any function of the body of man or other animals, and which does not achieve any of it's primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of any of its primary intended purposes.

U. Biological Specimens

83. Will biological specimens (including blood, urine and other human-derived samples) be used in this study?
☐ Yes Complete & Submit **Appendix 15**
☒ No

NOTE: If the response to Question 83 is YES, an application must be submitted to the Institutional Biosafety Committee (IBC). The IBC Applications may be located at http://www.research.psu.edu/orp/areas/biohazardous/applications/index.asp.

V. Other Biomedical Procedures – Diagnostic Radiation Procedures, Physical Activity, Diet Modifications

84. Will participants be asked to undergo **diagnostic** radiation procedures while enrolled in this study?
☐ Yes Complete & Submit **Appendix 16**
☒ No

85. Will participants be required to engage in or perform any form of physical activity?
☒ Yes Describe the nature and extent of the physical activity: **They will have to use a pen and paper, and they will use a keyboard and mouse.**
☐ No

86. Will any type of electrical equipment other than audio headphones be attached to the participants (e.g., EMG, EKG)?

☐ Yes Submit a letter describing the most recent safety check of the equipment with the supporting documents for this application.
☒ No

87. Will there be any diet modifications or restrictions?
☐ Yes Describe:
☒ No

W. Assurances

As the principal investigator on this research study, I assure that...

1. this application, if funded by an extramural source, accurately reflects all procedures involving human participants described in the grant proposal to the funding agency previously noted or an explanation is given for any differences.

2. I will obtain approval from the Institutional Review Board (IRB) before initiating any changes to the approved study, including changes in procedures, personnel, documents, instruments, etc., except where necessary to eliminate apparent immediate hazards to participants. In the latter instance, the IRB must be notified by the next workday.

3. I am familiar with and will comply with all pertinent institutional, local, state, and Federal regulations and policies. I will adhere to the policies and procedures described in Penn State's Federalwide Assurance with the Office for Human Research Protections as well as Federal regulations for the protection of human participants involved in research (45CFR46; 21CFR parts 50 & 56). Copies of these documents are available in the ORP upon request or on their website – http://www.research.psu.edu/orp/.

4. the information provided in this application reasonably summarizes the nature and extent of the proposed use of human participants.

5. I will notify the IRB within 5 business days regarding any significant adverse events that impact human participants.

6. all individuals listed on this form are competent and have been properly trained. I also assure that all individuals will complete the required training for the protection of human participants available on-line prior to contact with human participants.

7. any individual associated with or responsible for the design, the conduct, or the reporting of this research will comply with Penn State's Conflict of Interest Policy, RA-05.

_____ _____
Signature of Principal Investigator, REQUIRED Date

I hereby confirm that I have read this application and my signature denotes the completeness and accuracy of the information provided.

_____
PRINT Name of Faculty Advisor, REQUIRED IF PI IS A STUDENT

_____ _____
SIGNATURE of Faculty Advisor, REQUIRED IF PI IS A STUDENT Date

I hereby confirm that I have read this application and my signature denotes departmental/unit approval of this project. To the best of my knowledge, the information in the attached application relating to members of my department is correct.

The investigator(s) who are members of my department are qualified to perform the roles proposed for them in this application. Any novice researchers from my department will be supervised by qualified investigators.

PRINT Name of PI's Department/Unit Head, REQUIRED	
SIGNATURE of PI's Department/Unit Head, REQUIRED	Date

9 References

American Psychological Association. (2001). *The publication manual of the American psychological association* (5 ed.). New York: American Psychological Associations.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin Company.

Cozby, P. C. (2004). *Methods in behavioral research* (8th ed.). New York, NY: McGraw-Hill.

Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology*.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: Bradford Books/MIT Press.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling amplitude of movement. *Journal of Experimental Psychology, 47*(6), 381-391.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook*. Upper Saddle River: NJ: Prentice Hall/Pearson Education.

Kim, J. W. (2008). *Procedural skills: From learning to forgetting*. Unpublished doctoral dissertation, The Pennsylvania State University, University Park, PA.

Kim, J. W., Koubek, R. J., & Ritter, F. E. (2007). Investigation of procedural skills degradation from different modalities. In R. L. Lewis, T. A. Polk & J. E. Laird (Eds.), *Proceedings of the 8th International Conference on Cognitive Modeling* (pp. 255-260). Oxford, UK: Taylor & Francis/Psychology Press.

Kim, J. W., & Ritter, F. E. (2007). Automatically recording keystrokes in public clusters with RUI: Issues and sample answers. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (p. 1787). Austin, TX: Cognitive Science Society.

Kukreja, U., Stevenson, W. E., & Ritter, F. E. (2006). RUI: Recording user input from interfaces under Window and Mac OS X. *Behavior Research Methods, 38*(4), 656-659.

MacWhinney, B., St. James, J., Schunn, C., Li, P., & Schneider, W. (2001). STEP—A system for teaching experimental psychological using E-Prime. *Behavioral Research Methods, Instruments, & Computers, 33*(2), 287-296.

Mitchell, M. L., & Jolley, J. M. (2007). *Research design explained*. Belmont, CA: Thomson.

Montgomery, D. C. (2001). *Design and analysis of experiments* (5th ed.). New York, NY: John Wiley & Sons.

Nerb, J., Spada, H., & Ernst, A. M. (1997). A cognitive model of agents in a commons dilemma. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society* (pp. 560-565). Mahwah, NJ.: Erlbaum.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nielsen, J. (1994). Usability laboratories. *Behaviour & Information Technology, 13*(1-2), 3-8.

Ray, W. J. (2003). *Methods: Toward a science of behavior and experience* (7th ed.). Belmont, CA: Wadsworth/Thompson Learning.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 435-451.

Rempel, D., Willms, K., Anshel, J., Jaschinski, W., & Sheedy, J. (2007). The effects of visual display distance on eye accommodation, head posture, and vision and neck symptoms. *Human Factors, 49*(5), 830-838.

Ritter, F. E., & Larkin, J. H. (1994). Developing process models as summaries of HCI action sequences. *Human-Computer Interaction, 9*, 345-383.

Ritter, F. E., & Wood, A. B. (2005). Dismal: A spreadsheet for sequential data analysis and HCI experimentation. *Behavior Research Methods, 37*(1), 71-81.

Roediger, H. (2004). What should they be called? *APS Observer, 17*(4), 46-48.

Rosson, M. B., & Carroll, J. M. (2002). *Usability engineering: Scenario-based development of human-computer interaction*. San Francisco, CA: Morgan Kaufmann Publishers.

Salvucci, D. D. (in press). Rapid prototyping and evaluation of in-vehicle interfaces. *ACM Transactions on Computer-Human Interaction*.

Schoelles, M. J., & Gray, W. D. (2001). Argus: A suite of tools for research in complex cognition. *Behavior Research Methods, Instruments, & Computers, 33*(2), 130-140.

Stern, R. M., Ray, W. J., & Quigley, K. S. (2001). *Psychophysiological recording* (2nd ed.). New York, NY: Oxford University Press.

VanLehn, K. (2007). Getting out of order: Avoiding lesson effects through instruction. In F. E. Ritter, J. Nerb, T. O'Shea & E. Lehtinen (Eds.), *In order to learn: How the sequences of topics affect learning* (pp. 169-179). New York, NY: Oxford University Press.